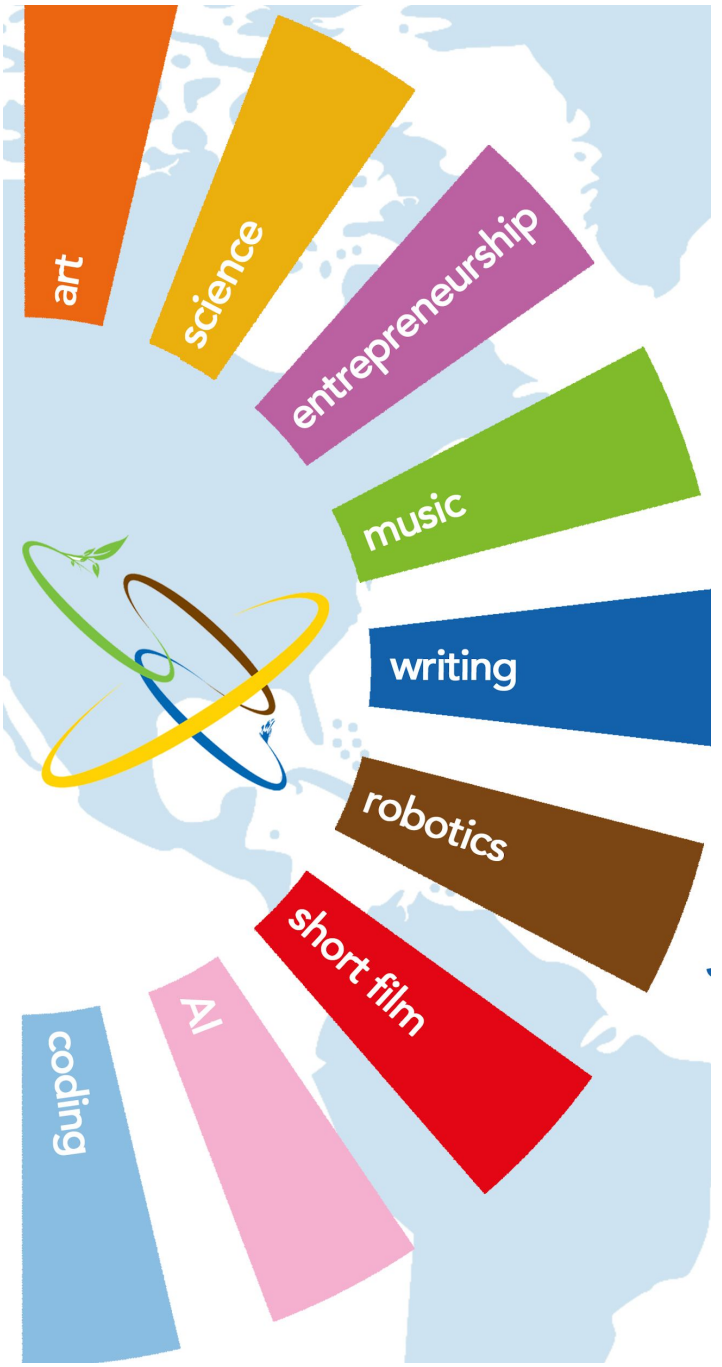
IJHSR International Journal of High School Research

October 2026 | Volume 8 | Issue 19
ijhsr.terrajournals.org
ISSN (Print) 2642-1046
ISSN (Online) 2642-1054

GENIUS OLYMPIAD

Let's build a better future together

International Project Fair focused on Sustainability and Environment for Grades 8-12



- Since 2011
- Hosted around 1400 participants in 2025 from 35+ states 70+ countries
- **Disciplines:** STEM, Coding, Robotics, AI, Speech, Entrepreneurship, Arts, Short Film, Music
- Applications start on December 1
- Application Deadline is March 1
- Finalists are announced by March 25
- Event is usually scheduled 2nd week of June
- Monday – Friday, includes a trip to Niagara Falls
- Hosted by large universities at Upstate New York
- Application Fee is $60/ project
- Participation Fee is $600/ person, w/ room and board
- Open buffet breakfast, lunch, and dinner
- Trip to Niagara Falls and boat tour is included
- **Instagram and Facebook @Geniusolympiad**
- For more information: **GENIUSOlympiad.org**
- Email: **info@geniusolympiad.org**

GENIUS Olympiad is organized by Terra Science and Education, a N.Y. based 501.c.3 non-profit organization dedicated for project-based learning

Table of Contents

October 2026| Volume 8| Issue 19

1 **Effect of Obesity-Related Genes on Survival Rates of Brain Tumor Patients**
Aarav Sehdev

13 **AI-based Detection of Fully Occluded Tomatoes Empowered by Pixel-based Lock-In-Amplifier and Synthetic Feature Injection**
Aiden Xu

22 **Extracting Influence: The Australian Mining Industry and National Climate Policy**
Andy Guan

32 **LGI1 in Epilepsy: Genetic Synaptopathies and Autoimmune Mechanisms**
Cara M. de Wesselow

39 **Anti-Price Gouging Laws and PPE Availability During the COVID-19 Pandemic in the United States**
Garvita Jain

46 **Parkinson's Disease: Genetic Factors, Emerging Therapies, and Their Implication for the Future**
Iris S. Lee

53 **A Comparison of Quantum Cryptography Protocols for Securing Internet of Things Devices**
Karlo N. Rendic

57 **What Drives the US Housing Affordability Crisis?**
Leonard Valckx

68 **Gut Microbiome and Sports Performance: The Symbiotic Relationship**
Megan Pan

75 **BioFETs with Magnetic Microbeads for Enhanced Biosensing and Early Disease Detection**
Melody Wang, Feng Gao, Hsueh-Chia Chang

86 **The Path to Green AI – Challenges and Solutions**
Nikhil Bhadra

92 **Diverse Biologic Modalities in Systemic Lupus Erythematosus (SLE) Clinical Trials**
Phoebe Hu

105 **Host Circadian Rhythm, Microbial Patterns, and Sleep Architecture**
Poorvi Kurundkar

114 **Is AI's Threat to Existential Unemployment Technologically Deterministic or Contingent on AI Governance?**
Rayan Moodley

121 **A Review of the Bilingual Advantage on the Onset of Dementia: The Knowledge Gap of Late L2 Acquisition**
Rio Imawaka

129 **Treating Leukemia Using CRISPR-Cas Genome Editing to Improve Delivery of Allogeneic and Autologous Stem Cells**
Rishabh A. Shah

138 **Enhancing Air Quality Forecasting by Predicting Volatile Organic Compounds Using Machine Learning**
Shreyas R. Rajalingari

144 **Applications of CRISPR/Cas9 in Modeling and Investigating Schizophrenia**
Siya Gupta

151 **A Functional and Socioeconomic Analysis of State-of-the-Art Prostheses Selection: Comparing Traditional, Myoelectric, and AI-Powered Devices**
Sheraya M. Amarasekara

162 **Structure, Dysfunction, and Therapeutic Targeting of the p53 Tumor Suppressor in Human Cancer**
Tanya Verma

172 **Supply Without Demand: Japan's Pension Reform and Elderly Male Labor Force Participation**
Terry S. Lo

182 **Laminar Flow: A Viable Focus Point for a Sustainable Aviation Future?**
Thomas Messeri

© 2026 Terra Science and Education

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Effect of Obesity-Associated Genes on Survival Rates of Brain Tumor Patients

Aarav Sehdev

Basis Phoenix, 13231 North 22nd St, Phoenix, AZ 85022, USA; aaravsehdev8@gmail.com

ABSTRACT: Brain tumors are a heterogeneous group of tumors, including low-grade gliomas (LGG) and glioblastoma multiforme (GBM). Obesity increases the risk of developing cancer, including brain tumors. However, there is conflicting evidence from epidemiological studies regarding the effect of obesity on brain tumors. We conducted a retrospective cohort study using publicly available databases to study the effect of obesity-associated genomic alterations on overall survival (OS) and disease-free survival (DFS) in brain tumor patients. We used the genome-wide association studies (GWAS) catalog to compile a list of genes associated with obesity. We hypothesized that brain tumor patients with alterations in obesity-associated genes (altered group) demonstrate better OS as compared to brain tumor patients without alterations (unaltered group). We found a significant improvement in the OS and a suggestion of improvement in the DFS in the altered group as compared to the unaltered group. Our study has several limitations due to its retrospective nature and the lack of availability of body mass index, thus inhibiting our ability to directly correlate clinical obesity status with genomic findings. Our analysis highlights the need for future preclinical and clinical research to elucidate the mechanisms of how obesity affects the survival outcomes in brain tumor patients.

KEYWORDS: Medical and Health Sciences, Oncology, Molecular Biology, Genetics, Bioinformatics.

Introduction

Brain tumors, including low-grade gliomas (LGG) and glioblastoma multiforme (GBM), are a heterogeneous group of tumors that arise from different types of glial or precursor cells, such as astrocytes, oligodendrocytes, and neural stem cells.[1] There are 24,820 estimated new cases of brain tumors, with 18,330 estimated deaths in the year 2025.[2] The 5-year relative survival is 33%.[2] These tumors are more common in males than in females due to hormonal and immunological differences.[3,4] The World Health Organization (WHO) classifies brain tumors into 4 grades, with grade 1 being benign (e.g., pilocytic astrocytoma) and grade 4 being the most malignant and aggressive brain tumor (e.g., glioblastoma multiforme). Grade 2 tumors are low-grade with usually a good prognosis after surgical excision, but with a 50% probability of anaplastic transformation (e.g., low-grade astrocytoma), whereas grade 3 tumors are more aggressive with a median survival of 24 months (e.g., anaplastic oligodendroglioma).[5] The development and progression of brain tumors is associated with environmental exposures such as ionizing radiation, carcinogens (such as N-nitroso compounds, pesticides, and air pollutants), and radiofrequency electromagnetic waves.[6] There is emerging evidence for an association between brain tumors and obesity.[7]

Obesity is now recognized as a major public health problem and is reaching epidemic proportions.[8] Obesity increases the risk of developing cancer, including brain tumors such as GBM and LGG; however, there is conflicting evidence from epidemiological studies.[9–11] Furthermore, emerging data suggest that taking glucagon-like peptide-1 receptor agonists lowers the risk of cancer, including meningioma, in obese or overweight individuals.[12] There is epidemiological and experimental evidence suggesting that obesity, measured by body mass index (BMI), is associated with a lower incidence and mortality from brain tumors – described as the "obesity paradox."[13–15] This is paradoxical because obesity is a known risk factor for multiple diseases such as diabetes, cardiovascular disease, and cancer. There are multiple plausible explanations for this paradox, including less aggressive cancer biology, a stronger immune system, and better tolerance to toxic treatments such as chemotherapy.[16–19] The molecular mechanisms underlying the outcomes in obese patients with brain tumors are not well understood. Preclinical and clinical evidence suggest that obesity promotes insulin resistance, resulting in increased levels of insulin and insulin-like growth factor (IGF), which act as potent mitogens by stimulating insulin receptor substrate (IRS).[20] In addition, adipose tissue can make a variety of pro-inflammatory cytokines and adipokines.[21] The altered levels of these cell-extrinsic factors regulate cellular metabolism through complex interactions with multiple signal transduction pathways.[20,21]

Genome-wide association studies (GWAS) have shown several genes to be associated with obesity.[22–24] Body mass index (BMI) was predominantly used as the proxy for obesity (BMI ≥ 30 kg/m2) in these studies.[25–27] Over the last 2 decades, more than 1000 single-nucleotide polymorphisms have been shown to have a significant association with obesity.[28] Additionally, several meta-analyses have further shown the strength of this association.[29–32] Therefore, it is plausible that the molecular alterations in the obesity-associated genes affect the survival of brain tumor patients.

Our hypothesis for this study is that brain tumor patients with alterations in obesity-associated genes demonstrate better

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.1

overall survival as compared to brain tumor patients without alterations in obesity-associated genes. Therefore, we conducted a retrospective cohort study with the primary aims of: i) describing the prevalence of obesity-associated genomic alterations in patients with brain tumors (LGG and GBM); ii) assessing the effect of obesity-associated genes on overall survival (OS) of patients with brain tumors; and iii) assessing the effect of obesity-associated genes on disease-free survival (DFS) of patients with brain tumors.

Methods

Patient Samples and Data Sources:

To test our hypothesis of correlation between obesity-associated genes and brain tumors, including LGG (astrocytoma grade II-III, anaplastic astrocytoma grade III, oligoastrocytoma grade III, anaplastic oligodendroglioma grade III, and anaplastic oligoastrocytoma grade III) and glioblastoma multiforme (GBM), we used the Cancer Genome Atlas (TCGA) data.[33,34] The TCGA data were accessed through cBioPortal, which is a widely available and easily usable online tool for clinical and genomics datasets for many different cancer types.[35–38] The Genomic Data Commons (GDC) data of brain tumor patients were obtained, which includes grade II to IV brain tumors, including astrocytoma, anaplastic astrocytoma, oligoastrocytoma, anaplastic oligodendroglioma, anaplastic oligoastrocytoma, and GBM.[39]

Using the TCGA data, brain tumor patients were analyzed for alterations in morbid obesity-associated genes. We used morbid obesity-associated genes instead of obesity-associated genes because of the stronger theoretical and statistical association, as well as the total number of genes showing association (53 unique genes for "morbid obesity," whereas 1122 genes for "obesity").

All available data, including gene mutation, copy number variations, and mRNA gene expression, were used to categorize the entire cohort into an altered group (those with genomic alterations) and an unaltered group (those without genomic alterations). Only non-synonymous mutations were used for gene mutation analyses. Synonymous mutations are not currently available in cBioPortal. The patients were classified into high and low mRNA expression using an RNA-seq z-score threshold of 2. The RNA-seq z-score is a normalization method that is used to determine how many standard deviations a gene's expression in a sample differs from the average expression across all samples. For instance, a z-score of 2 indicates the mean; a z-score higher than 2 indicates overexpression (high), and a z-score lower than 2 means underexpression (low). The final data lock date was December 6, 2025.

Primary Gene List: To identify obesity-associated genes, we used the Human GWAS Catalog. The National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) Catalog of human GWAS was accessed using the search term "morbid obesity" to identify the studies that have found significant gene association with morbid obesity.[40] The GWAS catalog provides both locus-level and mapped gene-level data in a tabular format. We used the mapped gene-level data. A total of 53 unique genes were identified, which were used for primary survival analysis and referred to as the "primary gene list" (Supplemental Table 1). Notably, the NHGRI-EBI Catalog is the largest known genome-wide association study database.

Alternate Gene Lists: To further assess the relationship between the obesity-associated genes and brain tumors, we also compiled 3 alternate gene lists of obesity-associated genes using other GWAS databases (Supplemental Table 2). These are described below:

i) The Centers for Disease Control and Prevention (CDC) Public Health Genomics and Precision Health Knowledge Base, version 10.0 (PHGKB).[41] PHGKB is based on a searchable online database of published scientific literature of genotype-phenotype associations. PHGKB is maintained by the Centers for Disease Control and Prevention (CDC). We found 15 genes showing association with obesity when we queried the PHGKB database for the search term "morbid obesity".

ii) Phenome-wide association study Web (PheWeb) is based on International Classification of Diseases (ICD) billing codes extracted from electronic health records (EHR) for genome-wide associations.[42] The PheWeb primarily has White British participants from the United Kingdom. We found 68 genes showing association with obesity when we queried the PheWeb database for the search term "obesity".

iii) A recent publication by Huang et al., where they analyzed 14 different tumor types using TCGA data to correlate molecular variances by BMI.[38] In this paper, they have reported 82 genes to be associated with higher BMI across 14 different tumor types.

Statistical Analysis:

Baseline characteristics between the altered group (brain tumor patients with genomic alterations in obesity-associated genes) and the unaltered group (brain tumor patients without genomic alterations in obesity-associated genes) were compared using a chi-square test for categorical variables and a Wilcoxon test for continuous variables. Q-value was calculated where relevant to control for the false discovery rate due to multiple hypothesis testing. Overall survival (OS) and disease-free survival (DFS) were estimated using the Kaplan–Meier method. The log-rank test was used to compare survival between the two groups (brain tumor patients with and without alterations in obesity-associated genes) using a P value < 0.05. A total of 10 patients who overlapped between the two groups were excluded from patient-level analysis. All statistical analyses were performed using cBioPortal or Microsoft® Excel for Mac Version 16.104.

Results

Study Participants:

A total of 530 LGG (diffuse glioma) and 539 GBM (glioblastoma multiforme) patients were in the study, of which 683 were in the altered and 386 were in the unaltered group. Within the altered group, there were significantly more patients with LGG compared to the unaltered group (406 vs. 124, respectively, $p < 10^{-10}$, $q < 10^{-10}$). Similarly, there were more

GBM patients in the altered group as compared to the unaltered group (277 vs. 262, respectively, $p < 10^{-10}$, $q < 10^{-10}$). The year of diagnosis for the altered group was between 1992 and 2013 versus 1989 and 2013 for the unaltered group ($p = 2.33e^{-4}$, $q = 4.95e^{-4}$). The patients in the altered group had less prior treatment compared to the unaltered group (3 vs 17, $p = 1.62e^{-5}$, $q = 4.25e^{-5}$). Altered and unaltered groups showed no significant difference between the mean mutation count (79.69 vs. 33.79, respectively, $p = 0.94$, $q = 0.96$) or mean fraction of genome altered (0.16 vs. 0.16, respectively, $p = 0.37$, $q = 0.53$). However, there were no significant differences between the two groups with respect to other demographic or clinical characteristics (Table 1).

Table 1: Characteristics of brain tumor patients with and without alterations in obesity genes. Importantly, there was no significant difference between the altered and unaltered groups in either mean mutation count or the mean fraction of the genome altered.

Clinical Attribute		Altered Group (N=683)	Unaltered Group (N=386)	Statistical Test	p-Value	q-Value
Diagnosis Age	Range	18-89	18-88	Wilcoxon Test	0.24	0.39
	Median	51	52			
	Mean	50.12	51.34			
Ethnicity Category	Not Hispanic or Latino	571	308	Chi-Squared Test	0.71	0.81
	NA	71	54			
	Hispanic or Latino	31	14			
Gender	Female	273	159	Chi-Squared test	0.50	0.62
	Male	398	213			
	NA	2	4			
Race	White	600	329	Chi-Squared Test	0.69	0.80
	Black or African American	42	25			
	NA	16	17			
	Asian	14	5			
	American Indian or Alaska Native	1	0			
Tumor Category	TCGA-LGG (Low-grade glioma/diffuse glioma)	406	124	Chi-Squared Test	$<10^{-10}$	$<10^{-10}$
	TCGA-GBM (Glioblastoma multiforme)	277	262			
Morphology	Glioblastoma, grade 4	270	252	Chi-Squared test	$<10^{-10}$	$<10^{-10}$
	NA	2	4			
	Diffuse astrocytoma, grade 2	41	24			
	Anaplastic astrocytoma, grade 3	106	25			
	Oligodendroglioma, grade 2	87	26			
	Oligoastrocytoma, grade 2	96	37			
	Anaplastic oligodendroglioma, grade 3	71	8			
Patient Primary Tumor site	Central nervous system (brain, spinal cord, and retina)	401	120	Chi-Squared test	$<10^{-10}$	$<10^{-10}$
	NA	2	4			
	Brain	270	252			
Year of Diagnosis	Minimum	1992	1989	Wilcoxon Test	$2.33e^{-4}$	$4.95e^{-4}$
	Maximum	2013	2013			
Mutation Count	Range	1-11177	1-131	Wilcoxon Test	0.94	0.96
	Median	30	31			
	Mean	79.69	33.79			
Fraction Genome Altered	Range	0-0.95	0-0.93	Wilcoxon Test	0.37	0.53
	Median	0.14	0.15			
	Mean	0.16	0.16			
Prior Treatment	False	665	355	Chi-Squared Test	$1.62e^{-5}$	$4.25e^{-5}$
	True	3	17			
	NA	5	4			
Prior Malignancy	False	400	136	Chi-Squared Test	0.96	0.96
	True	5	1			
	NA	268	239			

Genomic Alterations:

Analyzing the genomic alterations in obesity-associated genes, we found that 76.94% of LGG patients had alterations in these genes as compared to 51.33% of GBM patients. Patients with GBM showed a higher percentage of mutations (5.13% vs. 0.97%), amplifications (4.18% vs. 0.58%), and deep deletions (1.71% vs. 0.58%) as compared to LGG patients. However, a higher percentage of gene expression (62.4% vs. 30.23%) and multiple alterations (9.5% vs. 9.32%) were seen in LGG patients as compared to GBM patients (Table 2; Figure 1).

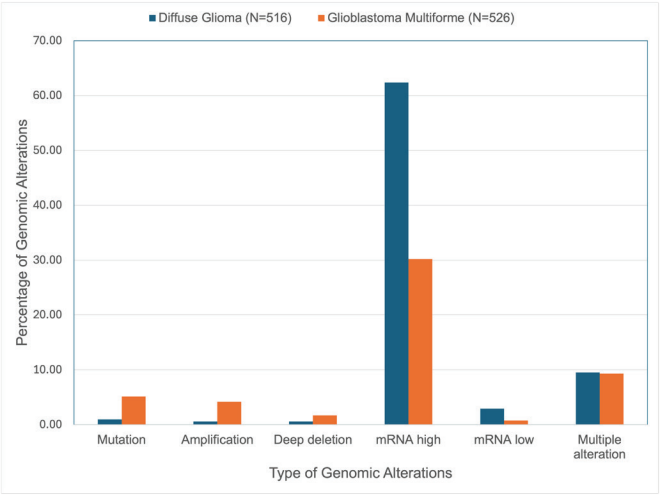


Figure 1: Cancer type and summary of genomic alterations in the altered vs. unaltered group. GBM patients had higher percentages of mutations, amplifications, and deep deletions, whereas LGG patients showed higher percentages of gene expression.

Table 2: Genomic alterations in obesity genes between diffuse glioma and glioblastoma multiforme. GBM patients showed more mutations, amplifications, and deep deletions, while LGG patients showed higher gene expression.

	Diffuse Glioma (N=516)	Glioblastoma Multiforme (N=526)
Mutation	0.97% (5 patients)	5.13% (27 patients)
Amplification	0.58% (3 patients)	4.18% (22 patients)
Deep deletion	0.58% (3 patients)	1.71% (9 patients)
mRNA high	62.4% (322 patients)	30.23% (159 patients)
mRNA low	2.91% (15 patients)	0.76% (4 patients)
Multiple alteration	9.5% (49 patients)	9.32% (49 patients)
Total	**76.94% (397 patients)**	**51.33% (270 patients)**

In the analysis of gene frequencies within the altered and unaltered group (as classified by the obesity-associated genes), we found that in the unaltered group the genes including CDKN2A (39.22% vs 27.29%, $p = 6.80e^{-5}$, $q = 2.36e^{-3}$), CDKN2B (38.18% vs 25.81%, $p = 3.02e^{-5}$, $q = 1.26e^{-3}$), UBA52P6 (37.92% vs 25.66%, $p = 3.89e^{-5}$, $q = 1.54e^{-3}$), CDKN2A-AS1 (37.40% vs 25.81%, $p = 9.40e^{-5}$, $q = 2.95e^{-3}$), CDKN2B-AS1 (35.32% vs 23.89%, $p = 8.99e^{-5}$, $q = 2.95e^{-3}$) showed significantly greater frequency as compared to the altered group (Table 3, Figure 2). Furthermore, we found that there were genes in the altered group, IDH1 (45.72% vs 31.17%, $p = 3.72e^{-6}$, $q = 4.95e^{-4}$), which showed significantly greater frequency in the altered group compared to the unaltered group (Table 3; Figure 2). Notably, the genes shown here (Table 3; Figure 2) are the genes with the highest frequency of alterations within these 2

DOI: 10.36838/IJHSR819.1

groups (not obesity-associated genes but rather any gene). In other words, after stratifying the cohort based on obesity-associated gene status, Figure 2 highlights the most frequently mutated genes in each group.

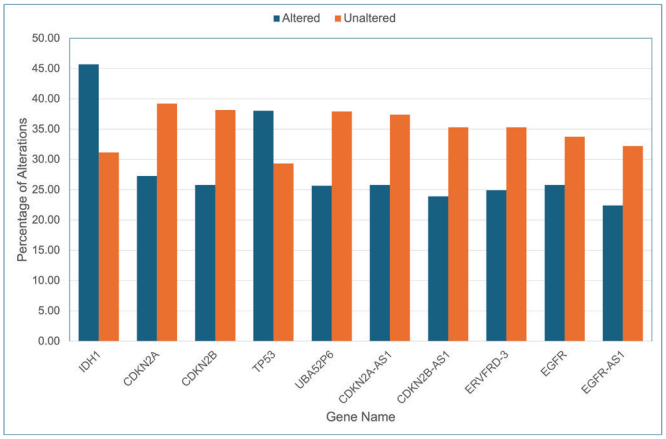


Figure 2: Genes with the highest frequency of variations in the altered and unaltered group. CDKN2A, CDKN2B, UBA52P6, CDKN2A-AS1, and CDKN2B-AS1 showed significantly greater frequency of modifications in the unaltered group, whereas IDH1 showed significantly greater frequency of modifications in the altered group.

Table 3: Genes with the highest frequency of alterations in the altered and unaltered group. CDKN2A, CDKN2B, UBA52P6, CDKN2A-AS1, and CDKN2B-AS1 were more frequently modified in the unaltered group, whereas IDH1 modifications were more common in the altered group.

Gene Name	Altered (N=678)	Unaltered (N=385)	P-value	Q-value
IDH1	45.72% (310)	31.17% (120)	$3.72e^{-6}$	$4.95e^{-4}$
CDKN2A	27.29% (185)	39.22% (151)	$6.80e^{-5}$	$2.36e^{-3}$
CDKN2B	25.81% (175)	38.18% (147)	$3.02e^{-5}$	$1.26e^{-3}$
TP53	38.05% (258)	29.35% (113)	$4.89e^{-3}$	0.09
UBA52P6	25.66% (174)	37.92% (146)	$3.89e^{-5}$	$1.54e^{-3}$
CDKN2A-AS1	25.81% (175)	37.40% (144)	$9.40e^{-5}$	$2.95e^{-3}$
CDKN2B-AS1	23.89% (162)	35.32% (136)	$8.99e^{-5}$	$2.95e^{-3}$
ERVFRD-3	24.93% (169)	35.32% (136)	$4.10e^{-4}$	0.01
EGFR	25.81% (175)	33.77% (130)	$7.28e^{-3}$	0.12
EGFR-AS1	22.42% (152)	32.21% (124)	$6.10e^{-4}$	0.02

Survival Analysis Using Primary Gene List:

The OS analysis included 1022 patients, 652 in the altered group and 370 in the unaltered group (Figure 3; Table 4). The median OS was significantly better in the altered group as compared to the unaltered group (33.21 *vs.* 20.37 months; HR = 0.70; 95% CI = 0.58-0.83; $p = 2.76e^{-5}$; q = $5.53e^{-5}$).

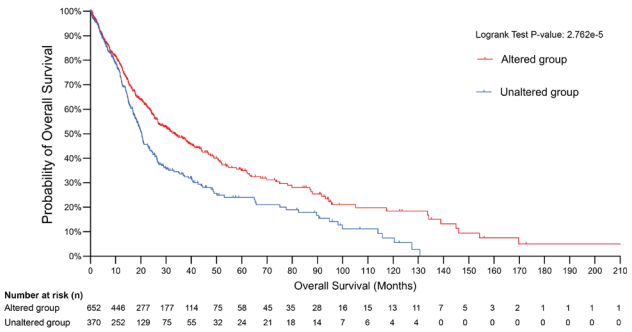


Figure 3: Overall survival (OS) in the altered vs. unaltered group. The median OS was significantly better in the altered group as compared to the unaltered group.

Table 4: Overall survival (OS) plot summary. Median OS was significantly higher in the altered group than in the unaltered group.

	Number of Cases, Total	Number of Events	Median Months Overall (95% CI)
Altered group	652	310	33.21 (26.91-41.10)
Unaltered group	370	226	20.37 (18.33-23.49)

The DFS analysis included 133 patients, 98 in the altered group and 35 in the unaltered group (Figure 4; Table 5). The median DFS was not statistically different between the two groups (median DFS not available; HR = 0.71; 95% CI = 0.27-1.87; $p = 0.45$; q = 0.45). Notably, Kaplan-Meier curves provide median DFS only when 50% of the cohort has experienced the event. As seen in Figure 4, this threshold has not been reached in our dataset for either group. Therefore, the median DFS cannot be estimated.

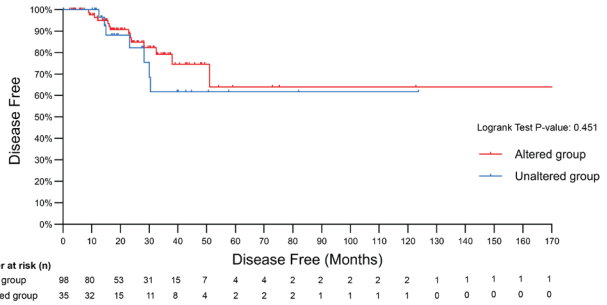


Figure 4: Disease-free survival (DFS) in the altered vs. unaltered group. The median DFS was not statistically different between the two groups.

Table 5: Disease-free survival (DFS) plot summary. Median DFS did not differ significantly between the two groups.

	Number of Cases, Total	Number of Events	Median Months Disease Free (95% CI)
Altered group	98	14	NA
Unaltered group	35	7	NA

Survival Analysis Using an Alternate Gene List:

i) *PHGKB*: The OS analysis included 1019 patients, 389 in the altered group and 630 in the unaltered group. The median OS was significantly better in the altered group as compared to the unaltered group (40.80 *vs.* 21.88 months; HR = 0.67; 95% CI = 0.57-0.80; $p = 1.79e^{-5}$; q = $3.58e^{-5}$).

The DFS analysis included 132 patients, 75 in the altered group and 57 in the unaltered group. The median DFS was not statistically different between the two groups (median DFS not available; HR = 1.08; 95% CI = 0.45-2.60; $p = 0.86$; q = 0.86).

ii) *PheWeb*: The OS analysis included 1023 patients, 716 in the altered group and 307 in the unaltered group. The median OS was significantly better in the altered group as compared to the unaltered group (32.42 *vs.* 19.84 months; HR = 0.63; 95% CI = 0.52-0.76; $p = 1.64e^{-7}$; q = $3.27e^{-7}$).

The DFS analysis included 134 patients, 113 in the altered group and 21 in the unaltered group. The median DFS was not statistically different between the two groups (median DFS not available; HR = 1.75; 95% CI = 0.54-0.57; p = 0.44; q = 0.44).

iii) Huang *et al.*: The OS analysis included 1032 patients, 959 in the altered group and 73 in the unaltered group. The median OS was significantly better in the altered group as compared to the unaltered group (27.2 *vs.* 17.64 months; HR = 0.57; 95% CI = 0.41-0.80; $p = 2.38e^{-5}$; q = $4.77e^{-5}$).

The DFS analysis could not be done due to not enough patients in the representative groups.

Discussion

In this large retrospective cohort study of 1069 patients with brain tumors, we report an observation of significant improvement in the OS and a suggestion of improvement in the DFS in patients with alterations in obesity-associated genes. Our results show that patients in the altered group have a significantly better prognosis as compared to the unaltered group. This has been described as the "obesity paradox" in brain tumors as well as in other cancers.[43–49] There are multiple plausible explanations for why obese patients have better outcomes. First, higher BMI in cancer patients could be reflective of less aggressive cancer biology, leading to better outcomes.[16] Second, most studies assess obesity using BMI, which can be misleading because elevated BMI may reflect greater muscle mass rather than increased adiposity.[17,18,43] Third, studies have shown a stronger immune system in obese patients, resulting in protective effects.[19] Lastly, patients with higher BMI may have greater nutrient reserves, potentially improving their tolerance to toxic treatments such as chemotherapy and radiation therapy.[16,43]

Our results showed better OS in the altered group compared to the unaltered group. One caveat in interpreting these results is the tumor type distribution between the altered and unaltered groups. Notably, the altered group included more LGG patients as compared to the unaltered group (406 vs. 124); however, this is unlikely to fully account for our findings, as the altered group also had a higher number of grade 3 brain tumor patients as compared to the unaltered group (177 vs. 32), partially balancing differences in tumor grade distribution. Therefore, we believe the distribution of LGG vs. GBM is probably not a significant factor by itself to account for the OS differences.

Strengths and Limitations :

There are several strengths of our study. First, we have used multiple gene lists that have shown association with obesity in different GWAS databases. Notably, there is not much overlap between these four different obesity-associated gene lists. For instance, the GWAS Catalog-based gene list (primary gene list) had only 7 genes in common with PHGKB, 3 genes in common with PheWeb, and 1 gene in common with Huang *et al.*[38] Finding significantly better OS results using different obesity-associated gene lists from different GWAS databases strengthens the accuracy of our findings. Second, our study is unique as we used publicly available databases to study the correlation between obesity-associated genes and brain tumors. To our knowledge, this correlation has not been previously reported. Additionally, this is the largest study examining the effect of obesity-associated genes and brain tumors in the existing literature. Third, we highlight the use of public databases to conduct research by high school students. Research opportunities for high-school and pre-medical undergraduate students remain limited, and many students drop out due to declining interest.[50] Increasing the availability of research opportunities can be a way to improve the interest of these students.[51] Using publicly available databases to conduct biomedical research can lead to high-impact research, as we demonstrate here.[52] Similar methods can be used to ask other important research questions by high-school and undergraduate students.

Our study has some limitations. First, we do not have the BMI data available on these patients, and therefore, we cannot be certain if the obesity-associated genes were more altered in the high BMI patients. However, there are no publicly available databases that have BMI and survival data available for brain tumor patients. Future research will need to be conducted with the collection of height and weight to confirm our findings. Second, this is a retrospective study and may have limitations such as selection bias and confounding. However, we have attempted to minimize selection bias by using a very large cohort and including all patients with brain tumors in the TCGA database. Third, there is potential bias from multiple hypothesis testing. However, we have calculated the false discovery rate and provided q-values where applicable. Fourth, it is possible that we have not been able to account for some of the post-translational effects of obesity on brain tumor patients. However, these studies will require laboratory experimentation and verification, which can be the next step in this area of research. Fifth, we used the GWAS catalog to identify obesity-associated genes based on mapped associations. However, the GWAS catalog primarily reports locus-level variant associations that are subsequently mapped to genes; this approach may introduce annotation bias and limit precision in gene-level analyses. Sixth, we recognize that more comprehensive analyses, such as multivariate approaches including the Cox proportional hazards model, could better account for potential confounding factors as well as pleiotropic effects of genes. However, the analyses presented in this manuscript rely on built-in tools available within cBioPortal, which currently do not support multivariate modeling. Nonetheless, our findings should be considered hypothesis-generating and may inform future investigations. Lastly, it is beyond the scope of the current work to conduct multivariate analyses, as this will require integrating multiple patient-level datasets and applying advanced statistical models, requiring expertise and resources not available for this student-led study.

Our study provides a holistic picture of the importance of obesity-associated genes in the prognosis of brain tumor patients. Given the limitations of our study (as described above) and the complex pathophysiology of obesity, our results are not clinically applicable as such. However, with further prospective research using clinical and biospecimen data collection, there will be a probable improvement in the outcomes of brain tumor patients. Therefore, we would like to propose the collection of such data in brain tumor patients. Additionally, we believe the advent of artificial intelligence (AI) and machine learning will tremendously improve the scope of healthcare research using publicly available databases by high school and undergraduate students.[53] Our study would not have been possible without the available tools in cBioPortal. However, cBioPortal only provides basic statistical tools and interactive exploration, but does not currently implement any built-in AI

DOI: 10.36838/IJHSR819.1

or machine learning tools. We are hopeful that future improvements in cBioPortal will make it possible to use more advanced AI analytical tools. Using these tools will make analyses of multi-omic datasets and identification of correlative markers more feasible, thereby improving patient outcomes.

Conclusion

In conclusion, in a large study using publicly available databases, we demonstrate that brain tumor patients with genomic alterations in obesity-associated genes have better overall survival than those without genomic alterations. Further preclinical and clinical studies are needed to elucidate the mechanisms of how obesity affects the survival outcomes in brain tumor patients. Surrogate markers for obesity (adiposity) such as BMI, fat mass index (FMI), and waist-to-hip ratio (WHR) should be collected during clinical studies in brain tumor patients.[54] This study highlights the importance of hypothesis-generating research using publicly available databases. This study can serve as a framework for other high-school and undergraduate students to develop and explore meaningful research questions. Lastly, enhancements in cBioPortal with built-in AI and machine learning tools will be immensely helpful to promote future complex research questions.

Acknowledgments

I would like to express my sincere gratitude to Dr. Yi Han Ng for her assistance in hypothesis formation, data collection, data interpretation, and manuscript writing. Additionally, I am deeply grateful to my science teachers, Ms. Kelsey Cooper and Dr. Thomas Carpenter, for teaching me the fundamentals of science and scientific writing.

References

1. Wiedmann, M. K. H.; Brunborg, C.; Di Ieva, A.; Lindemann, K.; Johannesen, T. B.; Vatten, L.; Helseth, E.; Zwart, J. A. The Impact of Body Mass Index and Height on the Risk for Glioblastoma and Other Glioma Subgroups: A Large Prospective Cohort Study. *Neuro-Oncol.* **2017**, *19* (7), 976–985. https://doi.org/10.1093/neuonc/now272
2. Surveillance Research Program, National Cancer Institute. Cancer Stat Facts: Brain and Other Nervous System Cancer. https://seer.cancer.gov/statfacts/html/brain.html (accessed 2025-11-28)
3. Yin, J.; Liu, G.; Zhang, Y.; Zhou, Y.; Pan, Y.; Zhang, Q.; Yu, R.; Gao, S. Gender Differences in Gliomas: From Epidemiological Trends to Changes at the Hormonal and Molecular Levels. *Cancer Lett.* **2024**, *598*, 217114. https://doi.org/10.1016/j.canlet.2024.217114
4. Yang, W.; Warrington, N. M.; Taylor, S. J.; Whitmire, P.; Carrasco, E.; Singleton, K. W.; Wu, N.; Lathia, J. D.; Berens, M. E.; Kim, A. H.; Barnholtz-Sloan, J. S.; Swanson, K. R.; Luo, J.; Rubin, J. B. Sex Differences in GBM Revealed by Analysis of Patient Imaging, Transcriptome, and Survival Data. *Sci. Transl. Med.* **2019**, *11* (473), eaao5253. https://doi.org/10.1126/scitranslmed.aao5253
5. Arulmani, D.; Manickam, D. Brain Tumors. *J. Stud. Res.* **2024**, *13* (2). https://doi.org/10.47611/jsrhs.v13i2.6694
6. Vienne-Jumeau, A.; Tafani, C.; Ricard, D. Environmental Risk Factors of Primary Brain Tumors: A Review. *Rev. Neurol. (Paris)* **2019**, *175* (10), 664–678. https://doi.org/10.1016/j.neurol.2019.08.004
7. Ahn, S.; Han, K.; Lee, J.-E.; Jeun, S.-S.; Park, Y.-M.; Yang, S. H. Associations of General and Abdominal Obesity with the Risk of Glioma Development. *Cancers* **2021**, *13* (12), 2859. https://doi.org/10.3390/cancers13122859
8. Chamarthi, V. S.; Daley, S. F. Comprehensive Assessment and Diagnosis of Metabolic and Biomechanical Complications in Obesity. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
9. Ahn, S.; Han, K.; Lee, J.-E.; Jeun, S.-S.; Park, Y.-M.; Yang, S. H. Associations of General and Abdominal Obesity with the Risk of Glioma Development. *Cancers* **2021**, *13* (12), 2859. https://doi.org/10.3390/cancers13122859
10. Sergentanis, T. N.; Tsivgoulis, G.; Perlepe, C.; Ntanasis-Stathopoulos, I.; Tzanninis, I.-G.; Sergentanis, I. N.; Psaltopoulou, T. Obesity and Risk for Brain/CNS Tumors, Gliomas and Meningiomas: A Meta-Analysis. *PLOS ONE* **2015**, *10* (9), e0136974. https://doi.org/10.1371/journal.pone.0136974
11. Zhang, D.; Chen, J.; Wang, J.; Gong, S.; Jin, H.; Sheng, P.; Qi, X.; Lv, L.; Dong, Y.; Hou, L. Body Mass Index and Risk of Brain Tumors: A Systematic Review and Dose-Response Meta-Analysis. *Eur. J. Clin. Nutr.* **2016**, *70* (7), 757–765. https://doi.org/10.1038/ejcn.2016.4
12. Dai, H.; Li, Y.; Lee, Y. A.; Lu, Y.; George, T. J.; Donahoo, W. T.; Lee, K. P.; Nakshatri, H.; Allen, J.; Guo, Y.; Sun, R. C.; Guo, J.; Bian, J. GLP-1 Receptor Agonists and Cancer Risk in Adults With Obesity. *JAMA Oncol.* **2025**, *11* (10), 1186–1193. https://doi.org/10.1001/jamaoncol.2025.2681
13. Obesity and Brain Tumors. In *The Link Between Obesity and Cancer*, Academic Press, 2023; pp 203–217. https://doi.org/10.1016/B978-0-323-90965-5.00011-8
14. Valente Aguiar, P.; Carvalho, B.; Vaz, R.; Linhares, P. Body Mass Index as an Independent Prognostic Factor in Glioblastoma. *Cancer Causes Control* **2021**, *32* (4), 327–336. https://doi.org/10.1007/s10552-020-01388-9
15. Wang, J.; Hao, Z.; Li, R.; Wu, W.; Huang, N.; Zhang, K.; Hao, S.; Feng, J.; Chu, J.; Ji, N. Association of Body Mass Index with Clinical Outcome of Primary WHO Grade 4 Glioma. *Front. Oncol.* **2024**, *14*. https://doi.org/10.3389/fonc.2024.1318785
16. Trestini, I.; Carbognin, L.; Bonaiuto, C.; Tortora, G.; Bria, E. The Obesity Paradox in Cancer: Clinical Insights and Perspectives. *Eat. Weight Disord. EWD* **2018**, *23* (2), 185–193. https://doi.org/10.1007/s40519-018-0489-y
17. Lennon, H.; Sperrin, M.; Badrick, E.; Renehan, A. G. The Obesity Paradox in Cancer: A Review. *Curr. Oncol. Rep.* **2016**, *18* (9), 56. https://doi.org/10.1007/s11912-016-0539-4
18. Park, Y.; Peterson, L. L.; Colditz, G. A. The Plausibility of the Obesity Paradox in Cancer-Point. *Cancer Res.* **2018**, *78* (8), 1898–1903. https://doi.org/10.1158/0008-5472.CAN-17-3043
19. Naik, A.; Monjazeb, A. M.; Decock, J. The Obesity Paradox in Cancer, Tumor Immunology, and Immunotherapy: Potential Therapeutic Implications in Triple Negative Breast Cancer. *Front. Immunol.* **2019**, *10*, 1940. https://doi.org/10.3389/fimmu.2019.01940
20. Kim, Y.-K.; Song, J. Metabolic Imbalance and Brain Tumors: The Interlinking Metabolic Pathways and Therapeutic Actions of Antidiabetic Drugs. *Pharmacol. Res.* **2025**, *215*, 107719. https://doi.org/10.1016/j.phrs.2025.107719
21. O'Flanagan, C. H.; Bowers, L. W.; Allott, E. H.; Hursting, S. D. Molecular and Metabolic Mechanisms Underlying the Obesity–Cancer Link. In *Energy Balance and Obesity*; Romieu, I., Dossus, L., Willett, W. C., Eds.; IARC Working Group Reports; International Agency for Research on Cancer: Lyon (FR), 2017.
22. Goodarzi, M. O. Genetics of Obesity: What Genetic Association Studies Have Taught Us about the Biology of Obesity and Its

Complications. *Lancet Diabetes Endocrinol.* **2018**, *6* (3), 223–236. https://doi.org/10.1016/S2213-8587(17)30200-0

23. Wang, K.; Li, W.-D.; Zhang, C. K.; Wang, Z.; Glessner, J. T.; Grant, S. F. A.; Zhao, H.; Hakonarson, H.; Price, R. A. A Genome-Wide Association Study on Obesity and Obesity-Related Traits. *PLoS ONE* **2011**, *6* (4), e18939. https://doi.org/10.1371/journal.pone.0018939
24. Krishnan, M.; Anwar, M. Y.; Justice, A. E.; Chittoor, G.; Chen, H.-H.; Roshani, R.; Smit, R. A. J.; Preuss, M. H.; Chami, N.; Hadad, B. S.; Parra, E. J.; Cruz, M.; Hui, Q.; Wilson, P. W. F.; Sun, Y. V.; Zhang, X.; Linchangco, G. V.; Kardia, S. L. R.; Faul, J. D.; Weir, D. R.; Bielak, L. F.; Highland, H. M.; Young, K. L.; Qi, B.; Wang, Y.; Fornage, M.; Haiman, C.; Cheng, I.; Peters, U.; Kooperberg, C.; Buyske, S.; McCormick, J. B.; Fisher-Hoch, S. P.; Lona-Durazo, F.; Peralta, J.; Gomez-Zamudio, J.; Rich, S. S.; Ferrier, K. R.; Lange, E. M.; Gignoux, C. G.; Kenny, E. E.; Wojcik, G. L.; Cho, K.; Gaziano, M. J.; Djousse, L.; Liu, S.; Vaidya, D.; Mutsert, R. de; Josyula, N. S.; Bauer, C. R.; Zhao, W.; Walker, R. W.; Smith, J. A.; Lange, L. A.; Meyer, M. C.; Liu, C.-T.; Yanek, L. R.; Lee, M.; Raffield, L. M.; Loos, R. J. F.; Gordon-Larsen, P.; Below, J. E.; North, K. E.; Graff, M. Genome-Wide Association Study Provides Novel Insight into the Genetic Architecture of Severe Obesity. medRxiv February 26, 2025, p 2025.02.25.25322870. https://doi.org/10.1101/2025.02.25.25322870
25. Krishnan, M.; Anwar, M. Y.; Justice, A. E.; Chittoor, G.; Chen, H.-H.; Roshani, R.; Scartozzi, A.; Dickerson, R. R.; Smit, R. A. J.; Preuss, M. H.; Chami, N.; Hadad, B. S.; Parra, E. J.; Cruz, M.; Hui, Q.; Wilson, P. W. F.; Sun, Y. V.; Zhang, X.; Linchangco, G. V.; Kardia, S. L. R.; Faul, J. D.; Weir, D. R.; Bielak, L. F.; Highland, H. M.; Young, K. L.; Qi, B.; Wang, Y.; Fornage, M.; Haiman, C.; Cheng, I.; Peters, U.; Kooperberg, C.; Buyske, S.; McCormick, J. B.; Fisher-Hoch, S. P.; Lona-Durazo, F.; Peralta, J.; Gomez-Zamudio, J.; Rich, S. S.; Ferrier, K. R.; Lange, E. M.; Gignoux, C. R.; Kenny, E. E.; Wojcik, G. L.; Cho, K.; Gaziano, M. J.; Djousse, L.; Liu, S.; Vaidya, D.; de Mutsert, R.; Josyula, N. S.; Bauer, C. R.; Zhao, W.; Walker, R. W.; Smith, J. A.; Lange, L. A.; Meyer, M. C.; Liu, C.-T.; Yanek, L. R.; Lee, M.; Raffield, L. M.; Loos, R. J. F.; Gordon-Larsen, P.; Below, J. E.; North, K. E.; Graff, M. Genome-Wide Association Study Provides Novel Insight into the Genetic Architecture of Severe Obesity. *PLOS Genet*. **2025**, *21* (9), e1011842. https://doi.org/10.1371/journal.pgen.1011842
26. Pulit, S. L.; Stoneman, C.; Morris, A. P.; Wood, A. R.; Glastonbury, C. A.; Tyrrell, J.; Yengo, L.; Ferreira, T.; Marouli, E.; Ji, Y.; Yang, J.; Jones, S.; Beaumont, R.; Croteau-Chonka, D. C.; Winkler, T. W.; Hattersley, A. T.; Loos, R. J. F.; Hirschhorn, J. N.; Visscher, P. M.; Frayling, T. M.; Yaghootkar, H.; Lindgren, C. M. Meta-Analysis of Genome-Wide Association Studies for Body Fat Distribution in 694 649 Individuals of European Ancestry. *Hum. Mol. Genet*. **2019**, *28* (1), 166–174. https://doi.org/10.1093/hmg/ddy327
27. Pei, Y.-F.; Zhang, L.; Liu, Y.; Li, J.; Shen, H.; Liu, Y.-Z.; Tian, Q.; He, H.; Wu, S.; Ran, S.; Han, Y.; Hai, R.; Lin, Y.; Zhu, J.; Zhu, X.-Z.; Papasian, C. J.; Deng, H.-W. Meta-Analysis of Genome-Wide Association Data Identifies Novel Susceptibility Loci for Obesity. *Hum. Mol. Genet*. **2014**, *23* (3), 820–830. https://doi.org/10.1093/hmg/ddt464
28. Ang, M. Y.; Takeuchi, F.; Kato, N. Deciphering the Genetic Landscape of Obesity: A Data-Driven Approach to Identifying Plausible Causal Genes and Therapeutic Targets. *J. Hum. Genet.* **2023**, *68* (12), 823–833. https://doi.org/10.1038/s10038-023-01189-3
29. Yengo, L.; Sidorenko, J.; Kemper, K. E.; Zheng, Z.; Wood, A. R.; Weedon, M. N.; Frayling, T. M.; Hirschhorn, J.; Yang, J.; Visscher, P. M.; GIANT Consortium. Meta-Analysis of Genome-Wide Association Studies for Height and Body Mass Index in ~700000 Individuals of European Ancestry. *Hum. Mol. Genet*. **2018**, *27* (20), 3641–3649. https://doi.org/10.1093/hmg/ddy271
30. Xu, M.; Li, M.; Zhang, Y.; Li, L.; Shen, Y.; Hu, G. Contributions of Clinical Obesity and Preclinical Obesity to the All-Cause Mortality Risk: Findings From the UK Biobank Cohort. *Diabetes Metab. Res. Rev.* **2025**, *41* (7), e70095. https://doi.org/10.1002/dmrr.70095
31. Smit, R. A. J.; Wade, K. H.; Hui, Q.; Arias, J. D.; Yin, X.; Christiansen, M. R.; Yengo, L.; Preuss, M. H.; Nakabuye, M.; Rocheleau, G.; Graham, S. E.; Buchanan, V. L.; Chittoor, G.; Graff, M.; Guindo-Martínez, M.; Lu, Y.; Marouli, E.; Sakaue, S.; Spracklen, C. N.; Vedantam, S.; Wilson, E. P.; Chen, S.-H.; Ferreira, T.; Ji, Y.; Karaderi, T.; Lüll, K.; Machado, M.; Malden, D. E.; Medina-Gomez, C.; Moore, A.; Rüeger, S.; Akiyama, M.; Allison, M. A.; Alvarez, M.; Andersen, M. K.; Appadurai, V.; Arbeeva, L.; Bartell, E.; Bhaskar, S.; Bielak, L. F.; Bis, J. C.; Bollepalli, S.; Bork-Jensen, J.; Bradfield, J. P.; Bradford, Y.; Brandl, C.; Braund, P. S.; Brody, J. A.; Broeckel, U.; Burgdorf, K. S.; Cade, B. E.; Cai, Q.; Camarda, S.; Campbell, A.; Cañadas-Garre, M.; Chai, J.-F.; Chesi, A.; Choi, S. H.; Christofidou, P.; Couture, C.; Cuellar-Partida, G.; Danning, R.; Degenhardt, F.; Delgado, G. E.; Delitala, A.; Demirkan, A.; Deng, X.; Dietl, A.; Dimitriou, M.; Dimitrov, L.; Dorajoo, R.; Eichelmann, F.; Eliasen, A. U.; Engmann, J. E.; Erdos, M. R.; Fairhurst-Hunter, Z.; Farmaki, A.-E.; Faul, J. D.; Fernandez-Lopez, J.-C.; Forer, L.; Frank, M.; Freitag-Wolf, S.; Fritsche, L. G.; Fuchsberger, C.; Galesloot, T. E.; Gao, Y.; Geller, F.; Giannakopoulou, O.; Giulianini, F.; Gjesing, A. P.; Goel, A.; Gordon, S. D.; Gorski, M.; Grove, J.; Guo, X.; Gustafsson, S.; Haessler, J.; Hansen, T. F.; Havulinna, A. S.; Haworth, S. J.; Heard-Costa, N.; Hemerich, D.; Highland, H. M.; Hindy, G.; Ho, Y.-L.; Hofer, E.; Holliday, E.; Horn, K.; Hornsby, W. E.; Hottenga, J.-J.; Huang, H.; Huang, J.; Huerta-Chagoya, A.; Huo, S.; Hwang, M. Y.; Hwu, C.-M.; Iha, H.; Ikeda, D. D.; Isono, M.; Jackson, A. U.; Jansen, I. E.; Jiang, Y.; Johansson, I.; Jonsson, A.; Jørgensen, T.; Kalafati, I. P.; Kanai, M.; Kanoni, S.; Kårhus, L. L.; Kasturiratne, A.; Katsuya, T.; Kawaguchi, T.; Kember, R. L.; Kentistou, K. A.; Kim, D.; Kim, H.-N.; Kim, Y. J.; Kleber, M. E.; Knol, M. J.; Kurbasic, A.; Lauzon, M.; Le, P.; Lea, R.; Lee, J.-Y.; Lee, W.-J.; Leonard, H. L.; Li, H.; Li, S. A.; Li, X.; Li, X.; Liang, J.; Lin, H.; Lin, K.; Liu, J.; Liu, X.; Lo, K. S.; Long, J.; Lores-Motta, L.; Luan, J.; Lyssenko, V.; Lyytikäinen, L.-P.; Mahajan, A.; Malik, M. Z.; Mamakou, V.; Mangino, M.; Manichaikul, A.; Marten, J.; Mattheisen, M.; McDaid, A. F.; Mei, Q.; Meiselbach, H.; Melendez, T. L.; Milaneschi, Y.; Miller, J. E.; Millwood, I. Y.; Mishra, P. P.; Mitchell, R. E.; Møllehave, L. T.; Mononen, N.; Mucha, S.; Munz, M.; Mykkänen, J.; Nakatochi, M.; Nardone, G. G.; Nelson, C. P.; Nethander, M.; Nho, C. W.; Nielsen, A. A.; Nolte, I. M.; Nongmaithem, S. S.; Noordam, R.; Ntalla, I.; Nutile, T.; Pandit, A.; Pauper, M.; Petersen, E. R. B.; Petersen, L. V.; Piluso, F.; Polašek, O.; Poveda, A.; Pyarajan, S.; Raffield, L. M.; Rakugi, H.; Ramirez, J.; Rasheed, A.; Raven, D.; Rayner, N. W.; Riveros, C.; Rohde, R.; Ruggiero, D.; Ruotsalainen, S. E.; Ryan, K. A.; Sabater-Lleal, M.; Santin, A.; Saxena, R.; Scholz, M.; Shen, B.; Shi, J.; Shin, J. H.; Sidore, C.; Sidorenko, J.; Sim, X.; Slieker, R. C.; Smith, A. V.; Smith, J. A.; Smyth, L. J.; Southam, L.; Steinthorsdottir, V.; Sun, L.; Takeuchi, F.; Taylor, K. D.; Tayo, B. O.; Tcheandjieu, C.; Terzikhan, N.; Tesolin, P.; Teumer, A.; Theusch, E.; Thompson, D. J.; Thorleifsson, G.; Timmers, P. R. H. J.; Trompet, S.; Turman, C.; Vaccargiu, S.; van der Laan, S. W.; van der Most, P. J.; van Klinken, J. B.; van Setten, J.; Verma, S. S.; Verweij, N.; Veturi, Y.; Wang, C. A.; Wang, C.; Wang, J.-S.; Wang, L.; Wang, Y. X.; Wang, Z.; Warren, H. R.; Bin Wei, W.; Wen, W.; Wheeler, W. A.; Wickremasinghe, A. R.; Wielscher, M.; Winsvold, B. S.; Wong, A.; Wuttke, M.; Xia, R.; Yamamoto, K.; Yang, J.; Yao, J.; Young, H.;

DOI: 10.36838/IJHSR819.1

Yousri, N. A.; Yu, L.; Zeng, L.; Zhang, W.; Zhang, X.; Zhao, J.-H.; Zhao, W.; Zhou, W.; Zimmermann, M. E.; Zoledziewska, M.; 't Hart, L. M.; Adair, L. S.; Adams, H. H. H.; Aguilar-Salinas, C. A.; Al-Mulla, F.; Arnett, D. K.; Asselbergs, F. W.; Åsvold, B. O.; Attia, J.; Banas, B.; Bandinelli, S.; Beilin, L. J.; Bennett, D. A.; Bergler, T.; Bharadwaj, D.; Biino, G.; Boerwinkle, E.; Böger, C. A.; Borja, J. B.; Bouchard, C.; Bowden, D. W.; Brandslund, I.; Brumpton, B.; Buring, J. E.; Caulfield, M. J.; Chambers, J. C.; Chandak, G. R.; Chanock, S. J.; Chaturvedi, N.; Ida Chen, Y.-D.; Chen, Z.; Cheng, C.-Y.; Cho, Y. S.; Christensen, K.; Christophersen, I. E.; Ciullo, M.; Cole, J. W.; Collins, F. S.; Concas, M. P.; Cooper, R. S.; Cruz, M.; Cucca, F.; Cutler, M. J.; Damrauer, S. M.; Dantoft, T. M.; de Borst, G. J.; de Geus, E. J. C.; de Groot, L. C. P. G. M.; De Jager, P. L.; de Kleijn, D. P. V.; de Silva, H. J.; Dedoussis, G. V.; den Hollander, A. I.; Du, S.; Easton, D. F.; Eckardt, K.-U.; Elders, P. J. M.; Eliassen, A. H.; Ellinor, P. T.; Elmståhl, S.; Erdmann, J.; Evans, M. K.; Fatkin, D.; Feenstra, B.; Feitosa, M. F.; Ferrucci, L.; Florez, J. C.; Ford, I.; Fornage, M.; Franke, A.; Franks, P. W.; Freedman, B. I.; Gieger, C.; Girotto, G.; Golightly, Y. M.; Gonzalez-Villalpando, C.; Gordon-Larsen, P.; Grallert, H.; Grant, S. F. A.; Grarup, N.; Griffiths, L.; Gudnason, V.; Haiman, C.; Hakonarson, H.; Hansen, T.; Hartman, C. A.; Hattersley, A. T.; Hayward, C.; Heid, I. M.; Heng, C.-K.; Hengstenberg, C.; Herzig, K.-H.; Hewitt, A. W.; Hishigaki, H.; Hougaard, D. M.; Hoyng, C. B.; Huang, P. L.; Huang, W.; Huang, W.-Y.; Huffman, J. E.; Hunt, S. C.; Hutri, N.; Hveem, K.; Hyppönen, E.; Iacono, W. G.; Ichihara, S.; Ikram, M. A.; Isasi, C. R.; Jarvelin, M.-R.; Jin, Z.-B.; Jöckel, K.-H.; Jonas, J. B.; Joshi, P. K.; Jousilahti, P.; Jukema, J. W.; Kähönen, M.; Kamatani, Y.; Kang, K. D.; Kaprio, J.; Kardia, S. L. R.; Karpe, F.; Kato, N.; Kavousi, M.; Kee, F.; Kessler, T.; Khera, A. V.; Khor, C. C.; Kiemeney, L. A. L. M.; Kim, B.-J.; Kim, E. K.; Kim, H.-L.; Kirchhof, P.; Kivimaki, M.; Koh, W.-P.; Koistinen, H. A.; Kokkinos, A.; Kooner, J. S.; Kooperberg, C.; Kovacs, P.; Kraaijeveld, A.; Kraft, P.; Krauss, R. M.; Kumari, M.; Kutalik, Z.; Laakso, M.; Lange, L. A.; Langenberg, C.; Launer, L. J.; Lee, H.; Lee, N. R.; Lehtimäki, T.; Lemaitre, R. N.; Li, H.; Li, L.; Lieb, W.; Lin, X.; Lind, L.; Linneberg, A.; Liu, C.-T.; Liu, J.; Loeffler, M.; London, B.; Lu, F.; Lubitz, S. A.; Mackey, D. A.; Magnusson, P. K. E.; Manson, J. E.; Marcus, G. M.; Marques Vidal, P.; Martin, N. G.; März, W.; Matsuda, F.; McCarthy, M. I.; McGarrah, R. W.; McGue, M.; McKnight, A. J.; Medland, S. E.; Mellström, D.; Metspalu, A.; Mitchell, B. D.; Mitchell, P.; Mook-Kanamori, D. O.; Mori, T. A.; Morris, A. D.; Mucci, L. A.; Munroe, P. B.; Nalls, M. A.; Nazarian, S.; Nelson, A. E.; Neville, M. J.; Newton-Cheh, C.; Nielsen, C. S.; Niinikoski, H.; Nikus, K.; Nöthen, M. M.; Ogunniyi, A.; Ohlsson, C.; Oldehinkel, A. J.; Orozco, L.; Pahkala, K.; Pajukanta, P.; Palmer, C. N. A.; Parra, E. J.; Pattaro, C.; Pedersen, O.; Pennell, C. E.; Penninx, B. W. J. H.; Perusse, L.; Peters, A.; Peyser, P. A.; Porteous, D. J.; Posthuma, D.; Power, C.; Pramstaller, P. P.; Province, M. A.; Psaty, B. M.; Qi, Q.; Qu, J.; Rader, D. J.; Raitakari, O. T.; Rallidis, L. S.; Rao, D. C.; Redline, S.; Reilly, D. F.; Reiner, A. P.; Rhee, S. Y.; Ridker, P. M.; Rienstra, M.; Ripatti, S.; Ritchie, M. D.; Rivadeneira, F.; Roden, D. M.; Rosendaal, F. R.; Rotter, J. I.; Rudan, I.; Rutters, F.; Ryu, S.; Sabanayagam, C.; Salako, B.; Saleheen, D.; Salomaa, V.; Samani, N. J.; Sanghera, D. K.; Sattar, N.; Schmidt, B.; Schmidt, H.; Schmidt, R.; Schulze, M. B.; Schunkert, H.; Scott, L. J.; Scott, R. J.; Sever, P.; Sheu, W. H. H.; Shoemaker, M. B.; Shu, X.-O.; Simonsick, E. M.; Sims, M.; Singleton, A. B.; Sinner, M. F.; Smith, J. G.; Snieder, H.; Spector, T. D.; Spedicati, B.; Stampfer, M. J.; Stark, K. J.; Strachan, D. P.; Tabara, Y.; Tai, E. S.; Tang, H.; Tardif, J.-C.; Thanaraj, T. A.; Tönjes, A.; Tuomi, T.; Tuomilehto, J.; Tusié-Luna, M.-T.; van Dam, R. M.; van der Harst, P.; Van der Velde, N.; van Duijn, C. M.; van Schoor, N. M.; Vitart, V.; Vohl, M.-C.; Völker, U.; Vollenweider, P.; Völzke, H.; Vrieze, S.; Wacher-Rodarte, N. H.; Walker, M.; Wander, G. S.; Wareham, N. J.; Watanabe, R. M.; Watkins, H.; Weir, D. R.; Werge, T. M.; Widen, E.; Willemsen, G.; Willett, W. C.; Wilson, J. F.; Wilson, P. W. F.; Wong, T. Y.; Woo, J.-T.; Wright, A. F.; Xu, H.; Yajnik, C. S.; Yang, J.; Yokota, M.; Yuan, J.-M.; Zeggini, E.; Zemel, B. S.; Zheng, W.; Zhu, X.; Zillikens, M. C.; Zonderman, A. B.; Zwart, J.-A.; Abecasis, G. R.; Assimes, T. L.; Auton, A.; Boehnke, M.; Chasman, D. I.; Esko, T.; Stefansson, K.; Lettre, G.; Lindgren, C. M.; Ng, M. C. Y.; O'Donnell, C. J.; Thorsteinsdottir, U.; Visscher, P. M.; Walters, R. G.; Winkler, T. W.; Wood, A. R.; Deloukas, P.; Frayling, T. M.; Justice, A. E.; Kilpeläinen, T. O.; Locke, A. E.; Mohlke, K. L.; North, K. E.; Okada, Y.; Willer, C. J.; Young, K. L.; Fatumo, S.; McCaffery, J. M.; Timpson, N. J.; Hirschhorn, J. N.; Sun, Y. V.; Berndt, S. I.; Loos, R. J. F. Polygenic Prediction of Body Mass Index and Obesity through the Life Course and across Ancestries. *Nat. Med.* **2025**, *31* (9), 3151–3168. https://doi.org/10.1038/s41591-025-03827-z

32. Larsson, S. C.; Burgess, S. Causal Role of High Body Mass Index in Multiple Chronic Diseases: A Systematic Review and Meta-Analysis of Mendelian Randomization Studies. *BMC Med.* **2021**, *19* (1), 320. https://doi.org/10.1186/s12916-021-02188-x
33. The Cancer Genome Atlas Research Network. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* **2015**, *372* (26), 2481–2498. https://doi.org/10.1056/NEJMoa1402121
34. Brennan, C. W.; Verhaak, R. G. W.; McKenna, A.; Campos, B.; Noushmehr, H.; Salama, S. R.; Zheng, S.; Chakravarty, D.; Sanborn, J. Z.; Berman, S. H.; Beroukhim, R.; Bernard, B.; Wu, C.-J.; Genovese, G.; Shmulevich, I.; Barnholtz-Sloan, J.; Zou, L.; Vegesna, R.; Shukla, S. A.; Ciriello, G.; Yung, W.; Zhang, W.; Sougnez, C.; Mikkelsen, T.; Aldape, K.; Bigner, D. D.; Van Meir, E. G.; Prados, M.; Sloan, A.; Black, K. L.; Eschbacher, J.; Finocchiaro, G.; Friedman, W.; Andrews, D. W.; Guha, A.; Iacocca, M.; O'Neill, B. P.; Foltz, G.; Myers, J.; Weisenberger, D. J.; Penny, R.; Kucherlapati, R.; Perou, C. M.; Hayes, D. N.; Gibbs, R.; Marra, M.; Mills, G. B.; Lander, E.; Spellman, P.; Wilson, R.; Sander, C.; Weinstein, J.; Meyerson, M.; Gabriel, S.; Laird, P. W.; Haussler, D.; Getz, G.; Chin, L. The Somatic Genomic Landscape of Glioblastoma. *Cell* **2013**, *155* (2), 462–477. https://doi.org/10.1016/j.cell.2013.09.034
35. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B. E.; Sumer, S. O.; Aksoy, B. A.; Jacobsen, A.; Byrne, C. J.; Heuer, M. L.; Larsson, E.; Antipin, Y.; Reva, B.; Goldberg, A. P.; Sander, C.; Schultz, N. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2012**, *2* (5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095
36. Gao, J.; Aksoy, B. A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S. O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; Cerami, E.; Sander, C.; Schultz, N. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **2013**, *6* (269), pl1. https://doi.org/10.1126/scisignal.2004088
37. de Bruijn, I.; Kundra, R.; Mastrogiacomo, B.; Tran, T. N.; Sikina, L.; Mazor, T.; Li, X.; Ochoa, A.; Zhao, G.; Lai, B.; Abeshouse, A.; Baiceanu, D.; Ciftci, E.; Dogrusoz, U.; Dufilie, A.; Erkoc, Z.; Garcia Lara, E.; Fu, Z.; Gross, B.; Haynes, C.; Heath, A.; Higgins, D.; Jagannathan, P.; Kalletla, K.; Kumari, P.; Lindsay, J.; Lisman, A.; Leenknegt, B.; Lukasse, P.; Madela, D.; Madupuri, R.; van Nierop, P.; Plantalech, O.; Quach, J.; Resnick, A. C.; Rodenburg, S. Y. A.; Satravada, B. A.; Schaeffer, F.; Sheridan, R.; Singh, J.; Sirohi, R.; Sumer, S. O.; van Hagen, S.; Wang, A.; Wilson, M.; Zhang, H.; Zhu, K.; Rusk, N.; Brown, S.; Lavery, J. A.; Panageas, K. S.; Rudolph, J. E.; LeNoue-Newton, M. L.; Warner, J. L.; Guo, X.; Hunter-Zinck, H.; Yu, T. V.; Pilai, S.; Nichols, C.; Gardos, S. M.; Philip, J.; AACR Project GENIE BPC Core Team, AACR Project

GENIE Consortium; Kehl, K. L.; Riely, G. J.; Schrag, D.; Lee, J.; Fiandalo, M. V.; Sweeney, S. M.; Pugh, T. J.; Sander, C.; Cerami, E.; Gao, J.; Schultz, N. Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res*. **2023**, *83* (23), 3861–3867. https://doi.org/10.1158/0008-5472.CAN-23-0816

38. Huang, F.; Xu, P.; Yue, Z.; Song, Y.; Hu, K.; Zhao, X.; Gao, M.; Chong, Z. Body Weight Correlates with Molecular Variances in Patients with Cancer. *Cancer Res*. **2024**, *84* (5), 757–770. https://doi.org/10.1158/0008-5472.CAN-23-1463
39. Heath, A. P.; Ferretti, V.; Agrawal, S.; An, M.; Angelakos, J. C.; Arya, R.; Bajari, R.; Baqar, B.; Barnowski, J. H. B.; Burt, J.; Catton, A.; Chan, B. F.; Chu, F.; Cullion, K.; Davidsen, T.; Do, P.-M.; Dompierre, C.; Ferguson, M. L.; Fitzsimons, M. S.; Ford, M.; Fukuma, M.; Gaheen, S.; Ganji, G. L.; Garcia, T. I.; George, S. S.; Gerhard, D. S.; Gerthoffert, F.; Gomez, F.; Han, K.; Hernandez, K. M.; Issac, B.; Jackson, R.; Jensen, M. A.; Joshi, S.; Kadam, A.; Khurana, A.; Kim, K. M. J.; Kraft, V. E.; Li, S.; Lichtenberg, T. M.; Lodato, J.; Lolla, L.; Martinov, P.; Mazzone, J. A.; Miller, D. P.; Miller, I.; Miller, J. S.; Miyauchi, K.; Murphy, M. W.; Nullet, T.; Ogwara, R. O.; Ortuño, F. M.; Pedrosa, J.; Pham, P. L.; Popov, M. Y.; Porter, J. J.; Powell, R.; Rademacher, K.; Reid, C. P.; Rich, S.; Rogel, B.; Sahni, H.; Savage, J. H.; Schmitt, K. A.; Simmons, T. J.; Sislow, J.; Spring, J.; Stein, L.; Sullivan, S.; Tang, Y.; Thiagarajan, M.; Troyer, H. D.; Wang, C.; Wang, Z.; West, B. L.; Wilmer, A.; Wilson, S.; Wu, K.; Wysocki, W. P.; Xiang, L.; Yamada, J. T.; Yang, L.; Yu, C.; Yung, C. K.; Zenklusen, J. C.; Zhang, J.; Zhang, Z.; Zhao, Y.; Zubair, A.; Staudt, L. M.; Grossman, R. L. The NCI Genomic Data Commons. *Nat. Genet.* **2021**, *53* (3), 257–262. https://doi.org/10.1038/s41588-021-00791-5
40. MacArthur, J.; Bowler, E.; Cerezo, M.; Gil, L.; Hall, P.; Hastings, E.; Junkins, H.; McMahon, A.; Milano, A.; Morales, J.; Pendlington, Z. M.; Welter, D.; Burdett, T.; Hindorff, L.; Flicek, P.; Cunningham, F.; Parkinson, H. The New NHGRI-EBI Catalog of Published Genome-Wide Association Studies (GWAS Catalog). *Nucleic Acids Res*. **2017**, *45* (D1), D896–D901. https://doi.org/10.1093/nar/gkw1133
41. Yu, W.; Gwinn, M.; Dotson, W. D.; Green, R. F.; Clyne, M.; Wulf, A.; Bowen, S.; Kolor, K.; Khoury, M. J. A Knowledge Base for Tracking the Impact of Genomics on Population Health. *Genet. Med. Off. J. Am. Coll. Med. Genet*. **2016**, *18* (12), 1312–1314. https://doi.org/10.1038/gim.2016.63
42. *Exploring and visualizing large-scale genetic associations by using PheWeb | Nature Genetics*. https://www.nature.com/articles/s41588-020-0622-5 (accessed 2025-11-30)
43. Cha, J.-Y.; Park, J.-S.; Hong, Y.-K.; Jeun, S.-S.; Ahn, S. Impact of Body Mass Index on Survival Outcome in Patients with Newly Diagnosed Glioblastoma: A Retrospective Single-Center Study. *Integr. Cancer Ther.* **2021**, *20*, 1534735421991233. https://doi.org/10.1177/1534735421991233
44. Schiffmann, J.; Karakiewicz, P. I.; Rink, M.; Manka, L.; Salomon, G.; Tilki, D.; Budäus, L.; Pompe, R.; Leyh-Bannurah, S.-R.; Haese, A.; Hammerer, P.; Huland, H.; Graefen, M.; Tennstedt, P. Obesity Paradox in Prostate Cancer: Increased Body Mass Index Was Associated with Decreased Risk of Metastases after Surgery in 13,667 Patients. *World J. Urol*. **2018**, *36* (7), 1067–1072. https://doi.org/10.1007/s00345-018-2240-8
45. Zhang, X.; Liu, Y.; Shao, H.; Zheng, X. Obesity Paradox in Lung Cancer Prognosis: Evolving Biological Insights and Clinical Implications. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **2017**, *12* (10), 1478–1488. https://doi.org/10.1016/j.jtho.2017.07.022
46. Ligibel, J. A.; Alfano, C. M.; Courneya, K. S.; Demark-Wahnefried, W.; Burger, R. A.; Chlebowski, R. T.; Fabian, C. J.; Gucalp, A.; Hershman, D. L.; Hudson, M. M.; Jones, L. W.; Kakarala, M.; Ness, K. K.; Merrill, J. K.; Wollins, D. S.; Hudis, C. A. American Society of Clinical Oncology Position Statement on Obesity and Cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2014**, *32* (31), 3568–3574. https://doi.org/10.1200/JCO.2014.58.4680
47. Lee, J.; Meyerhardt, J. A.; Giovannucci, E.; Jeon, J. Y. Association between Body Mass Index and Prognosis of Colorectal Cancer: A Meta-Analysis of Prospective Cohort Studies. *PLOS ONE* **2015**, *10* (3), e0120706. https://doi.org/10.1371/journal.pone.0120706
48. Chan, D. S. M.; Vieira, A. R.; Aune, D.; Bandera, E. V.; Greenwood, D. C.; McTiernan, A.; Navarro Rosenblatt, D.; Thune, I.; Vieira, R.; Norat, T. Body Mass Index and Survival in Women with Breast Cancer-Systematic Literature Review and Meta-Analysis of 82 Follow-up Studies. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2014**, *25* (10), 1901–1914. https://doi.org/10.1093/annonc/mdu042
49. Yuan, C.; Bao, Y.; Wu, C.; Kraft, P.; Ogino, S.; Ng, K.; Qian, Z. R.; Rubinson, D. A.; Stampfer, M. J.; Giovannucci, E. L.; Wolpin, B. M. Prediagnostic Body Mass Index and Pancreatic Cancer Survival. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol*. **2013**, *31* (33), 4229–4234. https://doi.org/10.1200/JCO.2013.51.7532
50. Lin, K. Y.; Parnami, S.; Fuhrel-Forbis, A.; Anspach, R. R.; Crawford, B.; De Vries, R. G. The Undergraduate Premedical Experience in the United States: A Critical Review. *Int. J. Med. Educ*. **2013**, *4*, 26–37. https://doi.org/10.5116/ijme.5103.a8d3
51. Estrada, M.; Burnett, M.; Campbell, A. G.; Campbell, P. B.; Denetclaw, W. F.; Gutiérrez, C. G.; Hurtado, S.; John, G. H.; Matsui, J.; McGee, R.; Okpodu, C. M.; Robinson, T. J.; Summers, M. F.; Werner-Washburne, M.; Zavala, M. Improving Underrepresented Minority Student Persistence in STEM. *CBE Life Sci. Educ*. **2016**, *15* (3), es5. https://doi.org/10.1187/cbe.16-01-0038
52. Hankey, W.; Zanghi, N.; Crow, M. M.; Dow, W. H.; Kratz, A.; Robinson, A. M.; Robinson, M. R.; Segarra, V. A. Using The Cancer Genome Atlas as an Inquiry Tool in the Undergraduate Classroom. *Front. Genet.* **2020**, *11*. https://doi.org/10.3389/fgene.2020.573992
53. Al-Nafjan, A.; Aljuhani, A.; Alshebel, A.; Alharbi, A.; Alshehri, A. Artificial Intelligence in Predictive Healthcare: A Systematic Review. *J. Clin. Med.* **2025**, *14* (19), 6752. https://doi.org/10.3390/jcm14196752
54. Khan, I.; Chong, M.; Le, A.; Mohammadi-Shemirani, P.; Morton, R.; Brinza, C.; Kiflen, M.; Narula, S.; Akhabir, L.; Mao, S.; Morrison, K.; Pigeyre, M.; Paré, G. Surrogate Adiposity Markers and Mortality. *JAMA Netw. Open* **2023**, *6* (9), e2334836. https://doi.org/10.1001/jamanetworkopen.2023.34836
55. Stelzer, G.; Rosen, N.; Plaschkes, I.; Zimmerman, S.; Twik, M.; Fishilevich, S.; Stein, T. I.; Nudel, R.; Lieder, I.; Mazor, Y.; Kaplan, S.; Dahary, D.; Warshawsky, D.; Guan-Golan, Y.; Kohn, A.; Rappaport, N.; Safran, M.; Lancet, D. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinforma*. **2016**, *54*, 1.30.1-1.30.33. https://doi.org/10.1002/cpbi.5
56. Centers for Disease Control and Prevention (CDC). Public Health Genomics and Precision Health Knowledge Base (PHGKB), 2024. https://phgkb.cdc.gov/PHGKB/phgHome.action
57. Gagliano Taliun, S. A.; VandeHaar, P.; Boughton, A. P.; Welch, R. P.; Taliun, D.; Schmidt, E. M.; Zhou, W.; Nielsen, J. B.; Willer, C. J.; Lee, S.; Fritsche, L. G.; Boehnke, M.; Abecasis, G. R. Exploring and Visualizing Large-Scale Genetic Associations by Using PheWeb. *Nat. Genet.* **2020**, *52* (6), 550–552. https://doi.org/10.1038/s41588-020-0622-5

Author

Aarav Sehdev is a sophomore at Basis Phoenix High School. He is very keen on advancing his understanding of biology and

DOI: 10.36838/IJHSR819.1

chemistry through scientific readings and experimentation. He has been elected as the Ambassador for Students Supporting Brain Tumor Research 2026-2027 and the Treasurer for the Science National Honor Society. He is vice-president and editorial board member for the Simplifications Project, which is focused on summarizing complex research papers to a high-school level. He enjoys playing tennis and guitar.

Supplemental Tables

Supplemental Table 1: Genome-wide association studies (GWAS) genes associated with morbid obesity.[55]

Gene Symbol	Gene Name	Function
FTO	Alpha-ketoglutarate-dependent dioxygenase	This gene is a nuclear protein of the AlkB-related non-heme iron and 2-oxoglutarate-dependent oxygenase superfamily, but the exact physiological function of this gene is not known. Other non-heme iron enzymes function to reverse alkylated DNA and RNA damage by oxidative demethylation. Studies in mice and humans indicate a role in the nervous and cardiovascular systems and a strong association with body mass index, obesity risk, and type 2 diabetes.
TCF7L2	Transcription Factor 7 Like 2	This gene encodes a high mobility group (HMG) box-containing transcription factor that plays a key role in the Wnt signaling pathway. The protein has been implicated in blood glucose homeostasis. Genetic variants of this gene are associated with increased risk of type 2 diabetes. Several transcript variants encoding multiple different isoforms have been found for this gene.
GIPR	Gastric Inhibitory Polypeptide Receptor	This gene encodes a G-protein-coupled receptor for gastric inhibitory polypeptide (GIP), which was originally identified as an activity in gut extracts that inhibited gastric acid secretion and gastrin release, but subsequently was demonstrated to stimulate insulin release in the presence of elevated glucose. Mice lacking this gene exhibit higher blood glucose levels with impaired initial insulin response after oral glucose load. Defects in this gene may thus contribute to the pathogenesis of diabetes.
LINC01875	Long intergenic Non-Protein coding RNA 1875	LINC01875 (Long Intergenic Non-Protein Coding RNA 1875) is an RNA Gene, and is affiliated with the lncRNA class.
TMEM18	Transmembrane Protein 18	Predicted to enable DNA binding activity. Involved in cell migration. Located in the nuclear membrane.
ETV5	ETS Variant Transcription Factor 5	Enables DNA-binding transcription activator activity, RNA polymerase II-specific, and RNA polymerase II transcription regulatory region sequence-specific DNA binding activity. Involved in cellular response to oxidative stress; negative regulation of transcription by RNA polymerase II; and positive regulation of transcription by RNA polymerase II. Located in the nucleoplasm.
PRDX4P1	Peroxiredoxin 4 Pseudogene 1	PRDX4P1 (Peroxiredoxin 4 Pseudogene 1) is a Pseudogene.
THAP12P9	THAP Domain Containing 12 Pseudogene 9	THAP12P9 (THAP Domain Containing 12 Pseudogene 9) is a Pseudogene.
LYPLAL1-AS1	LYPLAL1 Antisense RNA 1	LYPLAL1-AS1 (LYPLAL1 Antisense RNA 1) is an RNA Gene, and is affiliated with the lncRNA class. Diseases associated with LYPLAL1-AS1 include Developmental and Epileptic Encephalopathy 28.
SLC39A8	Solute Carrier Family 39 Member 8	This gene encodes a member of the SLC39 family of solute-carrier genes, which show structural characteristics of zinc transporters. The encoded protein is glycosylated and found in the plasma membrane and mitochondria, and functions in the cellular import of zinc at the onset of inflammation. It is also thought to be the primary transporter of the toxic cation cadmium, which is found in cigarette smoke. Multiple transcript variants encoding different isoforms have been found for this gene. Additional alternatively spliced transcript variants of this gene have been described, but their full-length nature is not known.
FAIM2	Fas Apoptotic Inhibitory Molecule 2	Predicted to enable calcium channel activity. Involved in the regulation of the neuron apoptotic process. Acts upstream of or within the negative regulation of the extrinsic apoptotic signaling pathway via death domain receptors. Located in the Golgi membrane and the membrane raft.
LINC02796	Long Intergenic Non-Protein Coding RNA 2796	LINC02796 (Long Intergenic Non-Protein Coding RNA 2796) is an RNA Gene, and is affiliated with the lncRNA class.
LINC03111	Long Intergenic Non-Protein Coding RNA 3111	LINC03111 (Long Intergenic Non-Protein Coding RNA 3111) is an RNA Gene, and is affiliated with the lncRNA class.
RNU4-17P	RNA, U4 Small Nuclear 17, Pseudogene	RNU4-17P (RNA, U4 Small Nuclear 17, Pseudogene) is a Pseudogene.
AKAP6	A-Kinase Anchoring Protein 6	The A-kinase anchor proteins (AKAPs) are a group of structurally diverse proteins that have the common function of binding to the regulatory subunit of protein kinase A (PKA) and confining the holoenzyme to discrete locations within the cell. This gene encodes a member of the AKAP family. The encoded protein is highly expressed in various brain regions and cardiac and skeletal muscle. It is specifically localized to the sarcoplasmic reticulum and nuclear membrane, and is involved in anchoring PKA to the nuclear membrane or sarcoplasmic reticulum.
NYAP2	Neuronal Tyrosine-Phosphorylated Phosphoinositide-3-Kinase Adapter 2	Predicted to be involved in neuron projection morphogenesis and phosphatidylinositol 3-kinase/protein kinase B signal transduction.
MIR-5702/5702	MicroRNA 5702	microRNAs (miRNAs) are short (20-24 nt) non-coding RNAs that are involved in post-transcriptional regulation of gene expression in multicellular organisms by affecting both the stability and translation of mRNAs. miRNAs are transcribed by RNA polymerase II as part of capped and polyadenylated primary transcripts (pri-miRNAs) that can be either protein-coding or non-coding. The primary transcript is cleaved by the Drosha ribonuclease III enzyme to produce an approximately 70-nt stem-loop precursor miRNA (pre-miRNA), which is further cleaved by the cytoplasmic Dicer ribonuclease to generate the mature miRNA and antisense miRNA star (miRNA*) products. The mature miRNA is incorporated into an RNA-induced silencing complex (RISC), which recognizes target mRNAs through imperfect base pairing with the miRNA and most commonly results in translational inhibition or destabilization of the target mRNA. The RefSeq represents the predicted microRNA stem-loop.
ADCY9	Adenylate Cyclase 9	Adenylate cyclase is a membrane-bound enzyme that catalyzes the formation of cyclic AMP from ATP. It is regulated by a family of G protein-coupled receptors, protein kinases, and calcium. The type 9 adenylyl cyclase is a widely distributed adenylyl cyclase, and it is stimulated by beta-adrenergic receptor activation but is insensitive to forskolin, calcium, and somatostatin.
MAP2K5	Mitogen-Activated Protein Kinase Kinase 5	The protein encoded by this gene is a dual specificity protein kinase that belongs to the MAP kinase kinase family. This kinase specifically interacts with and activates MAPK7/ERK5. This kinase itself can be phosphorylated and activated by MAP3K3/MEKK3, as well as by atypical protein kinase C isoforms (aPKCs). The signal cascade mediated by this kinase is involved in growth factor-stimulated cell proliferation and muscle cell differentiation. Three alternatively spliced transcript variants of this gene encoding distinct isoforms have been described.
LINC01741	Long Intergenic Non-Protein Coding RNA 1741	LINC01741 (Long Intergenic Non-Protein Coding RNA 1741) is an RNA Gene, and is affiliated with the lncRNA class.
SEC16B	SEC16 Homolog B, **Endoplasmic Reticulum Export Factor**	SEC16B is a mammalian homolog of S. cerevisiae Sec16 that is required for the organization of transitional endoplasmic reticulum (ER) sites and protein export
POM121C	**POM121 Transmembrane Nucleoporin C**	Predicted to enable nuclear localization sequence binding activity. Predicted to be a structural constituent of the nuclear pore. Predicted to be involved in RNA export from the nucleus and protein import into the nucleus. Located in the nuclear membrane and nucleoplasm.
DNAJB4	**DnaJ Heat Shock Protein Family (Hsp40) Member B4**	The protein encoded by this gene is a molecular chaperone, tumor suppressor, and member of the heat shock protein-40 family. The encoded protein binds the cell adhesion protein E-cadherin and targets it to the plasma membrane. This protein also binds incorrectly folded E-cadherin and targets it for endoplasmic reticulum-associated degradation. This gene is a strong tumor suppressor for colorectal carcinoma, and downregulation of it may serve as a good biomarker for predicting patient outcomes. Several transcript variants encoding different isoforms have been found for this gene.
GIPC2	**GIPC PDZ Domain Containing Family Member 2**	Enables identical protein binding activity. Located in the extracellular exosome.
TFAP2B	**Transcription Factor AP-2 Beta**	This gene encodes a member of the AP-2 family of transcription factors. AP-2 proteins form homo- or heterodimers with other AP-2 family members and bind specific DNA sequences. They are thought to stimulate cell proliferation and suppress terminal differentiation of specific cell types during embryonic development. Specific AP-2 family members differ in their expression patterns and binding affinity for different promoters. This protein functions as both a transcriptional activator and repressor. Mutations in this gene result in autosomal dominant Char syndrome, suggesting that this gene functions in the differentiation of neural crest cell derivatives.
MCM6	**Minichromosome Maintenance Complex Component 6**	The protein encoded by this gene is one of the highly conserved mini-chromosome maintenance proteins (MCM) that are essential for the initiation of eukaryotic genome replication. The hexameric protein complex formed by the MCM proteins is a key component of the pre-replication complex (pre_RC) and may be involved in the formation of replication forks and in the recruitment of other DNA replication-related proteins. The MCM complex, consisting of this protein and MCM2, 4, and 7 proteins, possesses DNA helicase activity and may act as a DNA unwinding enzyme. The phosphorylation of the complex by CDC2 kinase reduces the helicase activity, suggesting a role in the regulation of DNA replication. Single-nucleotide polymorphisms in the intron regions of this gene are associated with differential transcriptional activation of the promoter of the neighboring lactase gene and, thereby, influence lactose intolerance in early adulthood.
CAST	**Calpastatin**	The protein encoded by this gene is an endogenous calpain (calcium-dependent cysteine protease) inhibitor. It consists of an N-terminal domain L and four repetitive calpain-inhibition domains (domains 1-4), and it is involved in the proteolysis of the amyloid precursor protein. The calpain/calpastatin system is involved in numerous membrane fusion events, such as neural vesicle exocytosis and platelet and red-cell aggregation. The encoded protein is also thought to affect the expression levels of genes encoding structural or regulatory proteins. Alternatively spliced transcript variants encoding different isoforms have been described.
TDRG1	**Testis Development Related 1**	Predicted to be located in the cytoplasm. TDRG1 (Testis Development Related 1) is an RNA Gene, and is affiliated with the lncRNA class. Diseases associated with TDRG1 include Microvascular Complications of Diabetes 5 and Cervical Cancer.
LINC00951	**Long Intergenic Non-Protein Coding RNA 951**	LINC00951 (Long Intergenic Non-Protein Coding RNA 951) is an RNA Gene, and is affiliated with the lncRNA class. Diseases associated with LINC00951 include Immunodeficiency, Common Variable, 8, With Autoimmunity.
LINGO2	**Leucine Rich Repeat And Ig Domain Containing 2**	Predicted to act upstream of or within the positive regulation of synapse assembly. Predicted to be located in the membrane. Predicted to be active in several cellular components, including the extracellular space, glutamatergic synapse, and synaptic membrane.

SF3A2	**Splicing Factor 3a Subunit 2**	This gene encodes subunit 2 of the splicing factor 3a protein complex. The splicing factor 3a heterotrimer includes subunits 1, 2, and 3 and is necessary for the *in vitro* conversion of 15S U2 snRNP into an active 17S particle that performs pre-mRNA splicing. Subunit 2 interacts with subunit 1 through its amino-terminus, while the single zinc finger domain of subunit 2 plays a role in its binding to the 15S U2 snRNP. Subunit 2 may also function independently of its RNA splicing function as a microtubule-binding protein.
DTX2P1	**DTX2 Pseudogene 1**	DTX2P1 (DTX2 Pseudogene 1) is a Pseudogene.
UPK3BP1	**Uroplakin 3B Pseudogene 1**	UPK3BP1 (Uroplakin 3B Pseudogene 1) is a Pseudogene.
PMS2P11	**PMS1 Homolog 2, Mismatch Repair System Component Pseudogene 11**	Predicted to enable kinase activity. PMS2P11 (PMS1 Homolog 2, Mismatch Repair System Component Pseudogene 11) is a Pseudogene.
LINC01122	**Long Intergenic Non-Protein Coding RNA 1122**	LINC01122 (Long Intergenic Non-Protein Coding RNA 1122) is an RNA Gene, and is affiliated with the lncRNA class.
PPARG	**Peroxisome Proliferator-Activated Receptor Gamma**	This gene encodes a member of the peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors. PPARs form heterodimers with retinoid X receptors (RXRs), and these heterodimers regulate transcription of various genes. Three subtypes of PPARs are known: PPAR-alpha, PPAR-delta, and PPAR-gamma. The protein encoded by this gene is PPAR-gamma and is a regulator of adipocyte differentiation. Additionally, PPAR-gamma has been implicated in the pathology of numerous diseases, including obesity, diabetes, atherosclerosis, and cancer. Alternatively spliced transcript variants that encode different isoforms have been described.
RARB	**Retinoic Acid Receptor Beta**	This gene encodes retinoic acid receptor beta, a member of the thyroid-steroid hormone receptor superfamily of nuclear transcriptional regulators. This receptor localizes to the cytoplasm and to subnuclear compartments. It binds retinoic acid, the biologically active form of vitamin A, which mediates cellular signaling in embryonic morphogenesis, cell growth, and differentiation. It is thought that this protein limits the growth of many cell types by regulating gene expression. The gene was first identified in a hepatocellular carcinoma, where it flanks a hepatitis B virus integration site. Alternate promoter usage and differential splicing result in multiple transcript variants.
BDNF	**Brain-Derived Neurotrophic Factor**	This gene encodes a member of the nerve growth factor family of proteins. Alternative splicing results in multiple transcript variants, at least one of which encodes a preproprotein that is proteolytically processed to generate the mature protein. The binding of this protein to its cognate receptor promotes neuronal survival in the adult brain. Expression of this gene is reduced in Alzheimer's, Parkinson's, and Huntington's disease patients. This gene may play a role in the regulation of the stress response and in the biology of mood disorders.
BDNF-AS	**BDNF Antisense RNA**	BDNF-AS (BDNF Antisense RNA) is an RNA Gene, and is affiliated with the lncRNA class. Diseases associated with BDNF-AS include Glioblastoma and Multiple Myeloma.
RAPSN	**Receptor Associated Protein Of The Synapse**	This gene encodes a member of a family of proteins that are receptor-associated proteins of the synapse. The encoded protein contains a conserved cAMP-dependent protein kinase phosphorylation site and plays a critical role in clustering and anchoring nicotinic acetylcholine receptors at synaptic sites by linking the receptors to the underlying postsynaptic cytoskeleton, possibly by direct association with actin or spectrin. Mutations in this gene may play a role in postsynaptic congenital myasthenic syndromes. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene.
RPL35AP22	**Ribosomal Protein L35a Pseudogene 22**	RPL35AP22 (Ribosomal Protein L35a Pseudogene 22) is a Pseudogene.
ASTN2	**Astrotactin 2**	This gene encodes a protein that is expressed in the brain and may function in neuronal migration, based on functional studies of the related astrotactin 1 gene in human and mouse. A deletion at this locus has been associated with schizophrenia. Multiple transcript variants encoding different proteins have been found for this locus.
TMEM161B-DT	**TMEM161B Divergent Transcript**	Predicted to be a structural constituent of the ribosome. Predicted to be located in the ribosome. TMEM161B-DT (TMEM161B Divergent Transcript) is an RNA Gene, and is affiliated with the lncRNA class. Diseases associated with TMEM161B-DT include Glioma.
MIR9-2HG	**MIR9-2 Host Gene**	This is an evolutionarily conserved gene that produces alternatively spliced long non-coding RNAs that may be expressed predominantly in the brain and visual cortex. These transcripts may be involved in tumorigenesis, as depletion by siRNA suppressed glioma cell division. Transcripts may also bind to and regulate the activity of miR-411-5p and argonaut 2, thereby altering the expression of genes involved in tumor growth.
GLP1R	**Glucagon-Like Peptide 1 Receptor**	This gene encodes a 7-transmembrane protein that functions as a receptor for glucagon-like peptide 1 (GLP-1) hormone, which stimulates glucose-induced insulin secretion. This receptor, which functions at the cell surface, becomes internalized in response to GLP-1 and GLP-1 analogs, and it plays an important role in the signaling cascades leading to insulin secretion. It also displays neuroprotective effects in animal models. Polymorphisms in this gene are associated with diabetes. The protein is an important drug target for the treatment of type 2 diabetes and stroke. Alternative splicing of this gene results in multiple transcript variants.
DNAH8	**Dynein Axonemal Heavy Chain 8**	The protein encoded by this gene is a heavy chain of an axonemal dynein involved in sperm and respiratory cilia motility. Axonemal dyneins generate force through hydrolysis of ATP and binding to microtubules.
MIR1915HG	**MIR1915 Host Gene**	MIR1915HG (MIR1915 Host Gene) is an RNA gene and is affiliated with the lncRNA class.
SKIDA1	**SKI/DACH Domain Containing 1**	Predicted to enable DNA-binding transcription factor activity. Predicted to be involved in the regulation of transcription by RNA polymerase II.
AS3MT	**Arsenite Methyltransferase**	AS3MT catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to trivalent arsenical and may play a role in arsenic metabolism
BORCS7-ASMT	**BORCS7-ASMT Readthrough (NMD Candidate)**	This locus represents naturally occurring read-through transcription between the neighboring C10orf32 (chromosome 10 open reading frame 32) and AS3MT (arsenic, +3 oxidation state, methyltransferase) genes. The read-through transcript is a candidate for nonsense-mediated mRNA decay (NMD) and is therefore unlikely to produce a protein product.
SH2B1	**SH2B Adaptor Protein 1**	This gene encodes a member of the SH2-domain containing mediators family. The encoded protein mediates activation of various kinases and may function in cytokine and growth factor receptor signaling and cellular transformation. Alternatively spliced transcript variants have been described.
LINC02790	**Long Intergenic Non-Protein Coding RNA 2790**	LINC02790 (Long Intergenic Non-Protein Coding RNA 2790) is an RNA Gene, and is affiliated with the lncRNA class.
ZRANB3	**Zinc Finger RANBP2-Type Containing 3**	Enables ATP-dependent DNA/DNA annealing activity, DNA endonuclease activity, and K63-linked polyubiquitin modification-dependent protein binding activity. Involved in DNA repair, replication fork reversal, and response to UV. Located in the nuclear replication fork and nucleoplasm.

Supplemental Table 2: Gene lists from The CDC Public Health Genomics and Precision Health Knowledge Base, version 10.0 (PHGKB), and Phenome-wide association study Web (PheWeb), and Huang *et al.*[38,56,57]

PHGKB	**PheWeb**	**Huang *et al.***
TMEM18	*FTO*	PTEN
GNPDA2	*MC4R*	TXNIP
ETV5	*DEFB136*	TGFBR2
FAIM2	*TMEM18*	RUNX1
MTCH2	*SBK1*	KIF24
NEGR1	*DOCK11*	PCED1A
SH2B1	*SEMA3F*	LIPC
MAF	*SPANXB1*	YTHDC1
NPC1	*KCNJ15*	ACTB
ENPP1	*ZNF668*	NUP93
BDNF	*FAR1*	KMT2A
KCTD15	*ANXA1*	MYH9
PRRC2A	*PINX1*	CUL1
SEC16B	*MSRA*	TMCO4
PTER	*PRKG1*	FCRL6
	RGS18	MEGF8
	MCHR2	IKZF2
	SND1	HLA-A
	EDIL3	KLF5
	KCNJ6	SLC27A5
	OLFM4	DIAPH1
	C15orf48	TP53
	PPP3CA	ARAF
	TRAPPC9	AMER1
	FOXR2	PCBP1
	SEL1L	BRAF
	WNT7A	DYNC1I1
	CTSC	USP9X
	FGF1	CREB3
	ZNF722	ELF3
	XKR6	CTNNB1
	EPS8L2	ATP1B4
	BTBD3	DST
	PCDH11X	PTPN11
	LIPG	FAS
	BCL11B	ARID1A
	MRPS28	CD70
	LMO3	TMEM30A
	RSBN1	CIITA
	SMAD2	H1-2
	NDUFS4	DNAH10
	RGS7	HLA-B
	PCDH9	BIRC6
	GADL1	CDC42BPA
	PARK7	PZP
	GPC5	ERBB2
	MRLN	CUL3
	PSMG1	BAP1
	PLEKHM1	FAM20C
	ZNF716	ASPRV1
	ZNF236	TXNIP
	NALF1	KDM6A
	ZNF722	GTF2E1
	PTER	KRAS

DOI: 10.36838/IJHSR819.1

	FAM133A	TRIM71
	GGT1	IDE
	CFAP53	PPT2
	CD163	CDKN1B
	PABPC1L2B	ARHGEF17
	KLHL1	ADAM12
	LPP	KCTD20
	COX7C	USP7
	FAM220A	MFSD3
	MS4A10	SIRPD
	OLFM3	KDM6A
	ZFP37	DNAH14
	AIG1	RHCG
	ADGRG6	ZFP36L2
		TCF7L2
		BCL9
		CLCN5
		SMAD4
		PCDHGB4
		PTEN
		SGMS1
		CDKN2A
		RIPOR3
		CALHM4
		IDH1
		TNFSF14
		CFHR4
		CYSLTR2

Abbreviations: PHGKB, Public Health Genomics and Precision Health Knowledge Base; PheWeb, Phenome-wide association study Web.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

AI-based Detection of Fully Occluded Tomatoes Empowered by Pixel-based Lock-In-Amplifier and Synthetic Feature Injection

Aiden Xu

Winter Springs High School, 130 Tuskawilla Rd, Winter Springs, FL 32708, USA; goaidenxu@gmail.com
Mentor:

ABSTRACT: As robotic operations emerge in agricultural fields, the limited perception capability of robots in detecting fully occluded fruits remains a major obstacle to efficient harvesting, accurate yield prediction, and other field operations. It was previously shown that reflected electromagnetic wave signals are different between scenarios with and without fully occluded fruits. However, three challenges persist: the baseline scenario (i.e., the scenario without fully hidden fruits) is assumed to be known a priori; the signal difference is very small and embedded in a high level of noise; and no visible boundary can be seen with the unaided eye. In this paper, two image pre-processing techniques are proposed to aid classification and object detection tasks. First, a pixel-based lock-in amplifier (PLIA) is investigated to extract weak near-infrared (NIR) signals from experimental videos and output images. Then, the average RGB values in those images are embedded as synthetic features to enhance the performance of the classification and object detection models. Seventy-seven experiments were conducted in a mock tomato orchard setup, and NIR videos were collected, processed, and labeled. Two classification models (YOLOv11 and EfficientNet) and three object detection models (YOLOv11, YOLOv12, and RF-DETR) are trained, and their performances are compared. It is shown that all classification models achieve a precision above 70%, while the mAP50 values of the detection models exceed 50%. The proposed PLIA and synthetic feature injection (SFI) strategies improve the classification and object detection performance, as demonstrated in ablation tests, while the trained models are lightweight and easily deployable.

KEYWORDS: Applied AI Models, Fully Occluded Fruit Detection and Classification, Image/Video Pre-processing, Near Infrared, Precision Agriculture.

Introduction

Robotics has been utilized in many agricultural applications along with the rise of Agriculture 4.0.[1] Tasks such as sowing,[2] irrigation, monitoring,[3] yield prediction,[4] disease classification,[5] and harvesting,[6] among a myriad of other tasks, have seen varying degrees of automation to streamline and enhance their processes. High-value fruit crops, like apples, strawberries, and tomatoes,[7] in particular, have received significant research attention for the automation of their farming processes due to their labor-intensive nature. However, in order for these robots to be effective, they need unprecedented perception capabilities so that they can find fruits, either fully exposed, partially occluded, or fully obscured. Please note that "fully obscured" means that a fruit is completely behind leaves and that none of its parts can be observed by an RGB camera from any viewpoint. Finding them and in some cases classifying them are crucial for some robotic operations such as monitoring growth status,[8] managing disease conditions,[9] checking nutrient deficiency,[10] conducting yield prediction,[11] and harvesting.[12]

Computer vision has been widely used in detecting and classifying fruits based on RGB images.[13] Duong *et al.*[14] compared EfficientNet,[15] MixNet, and ConvNet, achieving an accuracy of 99.98%, 99.97%, and 94.52%, respectively, in the task of finding a variety of 25 different fruits and vegetables. As another example, in,[16] the detection of tomatoes achieved an average precision of 99.4% using the modified YOLOv3 model. Another study[17] focused on determining strawberry ripeness and found that a combination of YOLOv8 and Lightweight-Swin Transformer can reach an accuracy of 94.4%. Xiao *et al.*[18] utilized a vision transformer (Swin Transformer) and a multilayer perceptron to identify fruit ripeness and accomplished mAP50:95 values of 46.5%, 84.7%, and 83.9%, respectively. Additionally, Sa *et al.*[19] employed a YOLOv5 detection model and reached an mAP50:95 of 81.2%.

All the above-mentioned computer vision models focused on finding fully exposed or partially occluded fruits. These models detect and classify fruits based on the surface information of an image only. However, as seen in Figure 1, relying only on RGB images cannot detect fruits that are fully occluded. In many commercial farms, such as fresh market tomato farms, a significant amount of fruits are hidden inside dense leaves, and only counting the fruits that are visible to RGB cameras is not enough for many robotic operations, such as yield prediction and harvesting. Additionally, the approach of using cameras from different angles is not feasible, as leaves outside of the fruits are densely distributed.

In this study, a pixel-based lock-in-amplifier (PLIA) method and a synthetic feature injection (SFI) method are combined as an image/video processing method for the captured near infrared videos in seventy-seven experiments conducted in a mock tomato orchard setup. Then, two classification models (YOLOv11[20] and EfficientNet[15]) and three object detection models (YOLOv11, YOLOv12[20], and RF-DETR[21]) are custom-tuned and trained based on 875 processed images, trying

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.2

to classify whether or not there are tomatoes fully occluded by leaves, as well as using bounding boxes to locate where they are.



Figure 1: Two RGB images in a mock tomato orchard setup include tomatoes fully occluded by leaves, which are invisible to the unaided human eye. There are zero and two tomatoes hidden behind leaves in the left and right images, respectively.

The main contributions of this study are as follows:

• PLIA is developed to address the issue of reflected near-infrared (NIR) signals buried in a large magnitude of noise in captured images/videos. The modulation-demodulation-low pass filtering mechanism in the original lock-in amplifier (LIA) method[22] can help remove noise in all harmonic frequencies other than the DC component, which is the actual reflected NIR signal. Here, the LIA is applied to each pixel of a video along the temporal direction, so-called PLIA, and the result is an image instead of a scalar value, as shown in,[23] enabling the utilization of advanced convolutional neural network (CNN) and vision transformer models for later classification and detection tasks.

• The average RGB value of the PLIA processed image is injected into the image as a synthetic feature, a method similar to synthetic feature injection,[24] to enhance the performance of AI models in classification and object detection.

• A total of 875 NIR images, either collected from experiments or augmented via Roboflow®, are labeled and used in the CNN and vision transformer models' training, validation, and testing.

• CNN and vision transformer models were custom-tuned and trained to classify and detect fully hidden fruits given NIR images. For classification tasks, an EfficientNet and a YOLOv11 model were customized, and for detection tasks, a YOLO11, YOLOv12, and RF-DETR model were customized. Their performances are compared in terms of precision, recall, and F1 score for classification, and precision, recall, and mAP50 values for object detection, respectively.

It is worth noting that this study is a significant extension of the work presented in.[23] As shown in,[23] the RGB values of LIA-processed NIR images are different between scenarios with and without fully hidden fruits. However, the method proposed therein is not implementable because it assumes that the RGB value of scenarios without fully hidden fruits is known *a priori*, which is not the case under real farm conditions.

Related Work

Finding fruits that are fully occluded is a very challenging task, and there are only a few attempts in this area of interest. These methods can be largely grouped into two categories: mechanical approaches and electromagnetic wave approaches.

Several researchers have used mechanical approaches trying to expose the fruits hidden behind. For example, Jordi Gené-Mola *et al.*[25] tried to use leaf blowers to part the leaves so that they could see the fully hidden fruits behind. However, this approach was not highly effective because the leaves on most commercial tomato farms are tightly bound by strings, preventing the leaf blower from doing its intended purpose.

Reflected electromagnetic radiation, captured by a detector and converted from photons into electronic signals, can be used to determine whether fruits are fully blocked by leaves. (i) X-rays can penetrate leaves; however, the high hardware and software requirements, along with the associated cost, make them impractical for farm applications.[26] In addition, radiation would require safety gear to be worn, and may also damage fruits.[26] (ii) Although LiDAR operates in the NIR or short-wave infrared (SWIR) spectrum, existing products focus only on detecting and rendering surface information and are not capable of detecting objects behind leaves.[27] (iii) Millimeter radar may be a feasible approach, which has been proposed to detect hidden tomatoes with limited success in a one tomato plant setup; although this approach relies upon sophisticated hardware and signal processing software.[28] (iv) Mid-wave IR (MWIR) and long-wave IR (LWIR), such as thermal imagers, focus on finding objects that have temperature differences from their background. This approach may not work due to the fact that fruits and leaves do not have obvious temperature differences in general.[29] (v) Near infrared is shown to have relatively high transmissivity and low absorption and reflectivity for leaves and, at the same time, relatively high reflectivity and low transmissivity and absorption for tomato fruits.[23] Therefore, as shown in,[23] the reflected NIR signal is different between the scenarios with and without fully occluded fruits. However, as shown in Figure 2, it is still a very challenging task to detect if there are fully hidden fruits. (vi) Lastly, ultrasound, as a mechanical wave instead of an electromagnetic wave, has been proposed to tackle this challenging problem.[30] However, ultrasound signals have an insurmountable amount of noise from reflections off leaves and surroundings, and are also affected by local humidity and temperature.[30] Thus, it is not a feasible solution.

In brief, to the best of the author's knowledge, no effective approach has been proposed, including mechanical approaches, multi-angle vision systems, and electromagnetic approaches, that can effectively identify fruits fully hidden behind leaves.

Problem Definition

Based on the literature review, it is a challenging task to classify and detect whether there are fully occluded tomatoes. In this study, three questions will be answered via experiments in a mock tomato orchard setup, where the NIR emitter, the sensor detector, and the mock tomato orchard are placed in proximity.

• ***Question 1:*** Can we achieve a 70% or higher precision when using the proposed methods to classify whether there are fully hidden fruits behind leaves?

• ***Question 2:*** Can we achieve a 50% or higher mAP50 value when using the proposed methods to detect the number and locations of fully hidden fruits behind leaves?

• ***Question 3:*** Can we achieve a higher performance when incorporating the proposed PLIA and SFI methods?

The thresholds set in Question 1 and Question 2 are relatively lower than those in typical classification and object detection tasks. The reason is that the task of finding fully hidden fruits is extremely challenging. According to the initial study, in,[23] the RGB value difference in the raw NIR images between the cases with and without fully hidden fruits is very small, about 3% in average. Furthermore, as can be seen in Figure 2, there are no clearly visible, distinct boundaries of the fully hidden fruit.



Figure 2: NIR photons were detected by a cell phone camera, which generated the NIR images. There are two tomatoes fully hidden behind leaves in both the left and right images. The image on the left is the raw image, in which the fully hidden fruits cannot be seen. The center image shows the image after the PLIA preprocessing. The image on the far right is the corresponding RGB view of the other two images. Dotted lines on those images delineate the approximate locations of the two hidden tomatoes.

■ Overview of the Methods

The basic process of the method is shown in Figure 3. A group of thirty 940 nm diodes emits a pulse-width-modulated (PWM) modulated NIR signal towards the mock tomato orchard. Here, a special square PWM signal is generated in Arduino with 50% duty cycle and an amplitude between 0 and 1. Then the NIR wave propagates through the leaves and to the fruit(s) hidden behind if present. The NIR wave then reflects off the fully hidden fruit(s) and propagates back through the leaves. The photons are detected by a cellphone camera, which converts them digitally to an RGB video. This video then goes through the video pre-processing method of PLIA, which returns an image. Two filters are used in this step, including an average filter and a zero-phase low-pass filter (LPF). The SFI method is then applied to inject an additional visual clue. This processed image is then given to the custom-trained AI models, which will classify and detect fully hidden fruits. The result will determine the presence, number, and location of fully hidden fruits.

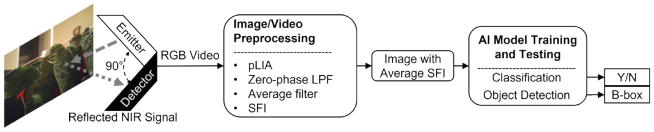


Figure 3: Overall description of the proposed method, including the experiment procedure, image/video pre-processing, and AI model training and testing.

■ Pre-Processing Methods

In the mock tomato orchard experiment, the PWM signal generation and video recording may have been slightly unsynchronized. Therefore, before pre-processing, the experiment videos should be synchronized with the reference PWM signal.

1. PLIA for video processing:

The lock-in amplifier (LIA) method has been widely used in extracting weak signals from a noisy environment.[22] Please note that the derivation here is similar to the original LIA,[22] but is custom-designed and applied to each pixel of an image frame along the temporal direction in an experiment video.

Let us define $A^I_{i,j}(k)$ as the RGB intensity of the photon that is reflected from fully occluded fruits and reaches the detector in frame k of a video at the pixel location of (i, j). Here, the RGB value means the sum of red, green, and blue values of a pixel in an image. $S^I_{i,j}(k)$ is $A^I_{i,j}(k)$ modulated with the PWM signal that is generated by the Arduino board as

$$S^I_{i,j}(k) = A^I_{i,j}(k)\left[\frac{1}{2} + \frac{2}{\pi}\sum_{n=1,3,5\ldots}^{\infty}\frac{1}{n}\sin(n\omega t_k)\right] \quad (1)$$

, in which the PWM signal is in its Fourier series expansion[31] and t_k is the time associated with frame k. Also, let's define $X^I_{i,j}(k) = S^I_{i,j}(k) + n_{i,j}(k)$ where $n_{i,j}$(k) is the noise captured by the detector and $X^I_{i,j}(k)$ is the photon signal detected by the camera at frame k and pixel location of (i, j) of a video.

The first step in PLIA is to demodulate $X^I_{i,j}(k)$ by multiplying it with the reference PWM signal, ri,j(k), in the form of a Fourier series for every pixel in an image frame of a video, along the temporal direction as

$$X_{i,j}(k)r_{i,j}(k) = [S^I_{i,j}(k) + n_{i,j}(k)] \times \left[\frac{1}{2} + \frac{2}{\pi}\sum_{m=1,3,5}^{\infty}\frac{1}{m}\sin(m\omega t_k)\right] \quad (2)$$

, the second line of which is the Fourier series of the reference PWM signal.[31] Then some trigonometric operations will be conducted, and a zero-phase LPF is applied to remove those non-DC terms. After that the Basel's equation[32] is used, the resultant value equals to $0.5A^I_{i,j}(k)$. Thus, we can extract the reflected photon signals from a high-noise environment.

This process will be repeated for every pixel of the experiment video. The achieved image is formed by averaging each achieved pixel value along the temporal direction, and the pixel value in the resultant image is

$$p_{i,j} = \sum_{k=1}^{N_f} 0.5A^I_{i,j}(k)/N_f \quad (3)$$

, in which N_f is the number of frames in this video. Since the reflected NIR photon is low, the converted RGB value $p_{i,j}$ is small. Therefore, to enhance the image quality, each pixel value of the image is scaled up so that the maximum pi,j in an image is near 255. The difference between a raw image and a PLIA processed image can be seen in Figure 2.

2. Synthetic feature injection:

As can be seen in Figure 2, the image obtained after processing an experiment video using PLIA is still not sufficiently distinct for reliable classification and object detection. As noted in,[23] the number of photons captured by the detector differs between scenarios with and without fully hidden fruits. Therefore, this characteristic is injected into the PLIA pre-processed NIR images.

DOI: 10.36838/IJHSR819.2

Three types of SFIs were tried based on the RGB value of an image after PLIA: average SFI (the average RGB value of an image), maximum SFI (the maximum RGB value of an image), and maximum difference SFI (the maximum RGB value - the minimum RGB value). Among them, the average SFI is selected because the failure rate of indicating whether there are hidden fruits is the lowest when tried on all the collected NIR images. Let's define the total RGB value of an image to be $V_i=R_i+B_i+G_i$, where i is the i^{th} pixel in the image. Then the average SFI p_{sfi} is calculated as all V_i summed together and then divided by N, where N is the total number of pixels in an image. Since p_{sfi} in all the experiments fall within a certain range, this value is adjusted to be a[p_{sfi} - p_{base}] to better scale the subsequently constructed black square box. Here, p_{base} and a are tuned based on the experiment data. This value is proportional to the side length of the black square placed at an appropriate point in the NIR image, and in this case, the top right corner.

One resultant image after these two image/video processing methods is shown in Figure 4. Since the average RGB value is proportional to the number of fully hidden fruits present, the SFI method gives additional artificial visual data injected directly into the images, thus improving the accuracy metrics of the AI models.

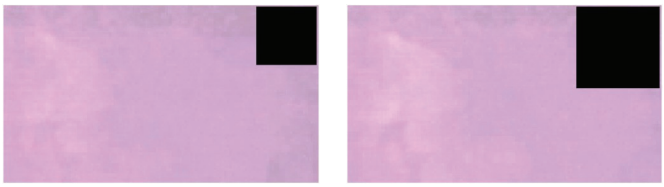


Figure 4: Images after the application of SFI. The average total RGB value is represented by the side length of the black square in the top right corner. This feeds additional visual information into the AI models, improving their performance. An image with one fruit is shown on the left, and an image with two fruits is shown on the right. There is a noticeable difference in the side lengths of the squares between the two.

3. Image/video pre-processing:

With the PLIA and SFI methods, the pre-processing steps are shown in Figure 5. Each video is downsized from 1920x1080 to 480x270 due to memory limits, as well as the computational cost consideration of the low-pass filter used. These downsized videos are converted to a 4-dimensional matrix for processing. Then, the signals are aligned to the reference PWM signal and then truncated to 4 periods. To reduce the memory needs and computational cost, only one period of the video, about 6 seconds, is used. The signals are passed to the PLIA method, which applies the LIA method to each pixel in the video and returns an RGB image. Each pixel is then multiplied by 255 over its maximum total RGB value. Then, they go through the SFI method, which injects additional visual information into the image in the form of a black square. The side length of this square is equal to 10 times the quantity of the average total RGB value minus 57, i.e., p_{base} =57 and a=10. One example image is shown in Figure 4.

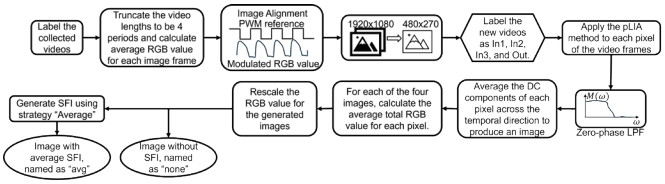


Figure 5: Procedure for image/video preprocessing. The outputs include eight pictures, with one set of four having 0, 1, 2, and 3 fruits, and the other set of four each with SFI.

Experiment and Video Collections

1. Testbed apparatus and setup:

The same apparatus as the one in[23] is used in this study, with a few modifications. The detailed construction of such a low-cost, mock tomato orchard and the emitter/detector therein can be found in.[23] To increase the number of photons captured by the camera, and meanwhile to maintain the low cost, thirty 940nm NIR diodes are integrated in the emitter, and all the circuits are on printed circuit boards to reduce the number of wires needed and save space. A cell phone camera is used to detect the reflected NIR wave.

2. Video collection:

The experiment procedure is shown in Figure 6. In each experiment, the leaves are taped to strings that go across the front to simulate the leaf cover. Up to three tomatoes are then placed just behind the leaf cover, with the tomatoes and leaves adjacent to each other and at a total distance of roughly 12 cm from the emitter. When the experiment starts, the emitter emits an NIR signal modulated with PWM. The PWM signal has a period of 6 seconds at a duty cycle of 50%. The camera detector (Motorola Moto G Pure®) detects the reflected NIR photons at a frame rate of 30 frames per second and digitally converts them into an RGB video. This experiment will repeat four cases with 0, 1, 2, and 3 fully hidden tomatoes, respectively.

A total of 77 valid experiments were conducted mainly in the evening to reduce the noise level. These generated images are processed with PLIA and SFI, then labeled using labelImg®. The initial YOLO dataset was copied, and that copy was reformatted to be compatible with EfficientNet and RF-DETR using the Roboflow® platform. The datasets are also augmented in Roboflow® using vertical and horizontal flips, as well as 90-degree rotations.

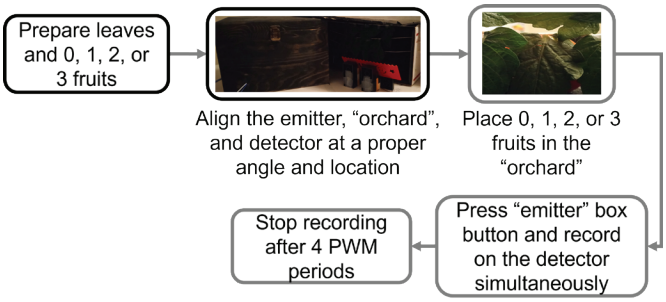


Figure 6: Procedure for collecting experimental data. The output includes four videos, with 0, 1, 2, and 3 fruits in them, respectively. These videos are passed to the image preprocessing procedure as described in Figure 5.

Classification and Object Detection Pipelines

In classification, as shown in Figure 7, there will be two classes: no fruit and fruits. For EfficientNet, hyperparameters including the model, image size, batch size, dropout threshold, learning rate, and the seed are carefully tuned; while for YOLOv11, hyperparameters include the model, image size, batch size, cls value, dropout threshold, learning rate, and label smoothing value.

As shown in Figure 7, the classification pipeline starts with setting up the dataset and output directories, and then the GPU is activated for YOLOv11. After that, a specific classification model is retrieved, and two classes are defined. Hyperparameters are tuned, and for EfficientNet, a random seed is picked. The model is then trained based on processed images. If the precision is less than 0.7, the model will be retrained using new configured hyperparameters. If the precision is equal to or higher than 0.7, the result will be compared with the saved best training result. If the current model is not better than the saved one, hyperparameters will be adjusted, and the classification model will be retrained. Otherwise, the model with the best precision value will be saved.

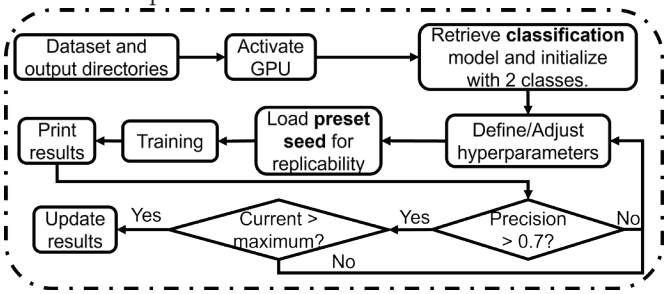


Figure 7: Classification pipeline for determining if there are fruits fully hidden behind leaves. Two AI models are trained: YOLOv11 and EfficientNet.

For object detection models, as shown in Figure 8, the base model is initialized with one class. For YOLOv11 and YOLOv12, tuned hyperparameters include the model (nano, small, medium, and large), image size, batch size, cls value, dropout threshold, learning rate, and label smoothing. For RF-DETR, hyperparameters include batch size, learning rate, backbone learning rate, resolution, and weight decay. Those hyperparameters are discussed in the results sections.

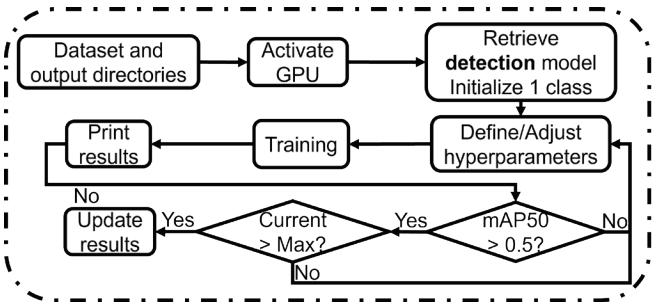


Figure 8: Object detection pipeline used to train a model to decide the number and locations of possible fruits fully hidden behind leaves. Three AI models are trained here: YOLOv11, YOLOv12, and RF-DETR.

It is worth noting that the CPU is used to train the EfficientNet model because of the need to strictly control the random seed value for the replicability of results. GPU is used to train all four other AI models.

Results and Discussion

To answer the three questions, four AI models, PLIA, and SFI are teamed in five different combinations, and their performance is compared. Among them, EfficientNet and YOLOv11 are used in classification, while YOLOv11, YOLOv12, and RF-DETR are compared in terms of object detection performance.

The first two sections will discuss the best results in classification and object detection, respectively. The last section will discuss the ablation tests, demonstrating whether PLIA and SFI improve performance.

All training, validation, and testing are conducted on a laptop computer. Its CPU (for EfficientNet) is a 13th Gen Intel(R) CoreTM i7-13700HX (2.10 GHz) with a RAM of 16GB, and the GPU used in YOLOv11, YOLOv12, and RF-DETR is the NVIDIA GeForce RTX 4060. The Python version used is 3.10.19 for EfficientNet and 3.11.14 for YOLOv11, YOLOv12, and RF-DETR.

To achieve optimal classification and/or object detection performance, the AI models were trained across a wide range of hyperparameter settings. For each hyperparameter setting, performance metrics such as F1 score, mAP50, and precision are recorded. Through extensive trials, we observed that model performance is mainly affected by three hyperparameters: learning rate, image size, and batch size. Therefore, based on this observation, a total of 45 hyperparameter settings are tried for two classification and two object detection models. The only exception is the RF-DETR model, which requires a long training time for each hyperparameter setting. Therefore, only a limited number of settings were evaluated for the RF-DETR model. The 45 combinations of the hyperparameter settings are: learning rate (0.00005, 0.0001, 0.0002), image size (224, 288, 320, 480, 640), and batch size (8, 16, 24).

The images in the dataset are split into three groups: training (70%), validation (20%), and testing (10%). All images from the same experiment are assigned to the same dataset partition. The dataset is split at the experiment level, rather than at the image level.

It is worth noting that approximate tomato placement is determined beforehand via ordinary RGB pictures taken of the apparatus, as shown in Figure 2. This information is utilized to determine the ground-truth bounding boxes by noting the approximate locations of each of the tomatoes, the number of tomatoes in that specific experiment, and the size of the tomatoes.

1. Classification results

:Two AI models, EfficientNet and YOLOv11, enhanced by PLIA and SFI, are customized for classification tasks to determine whether there are tomatoes fully occluded by leaves in the mock tomato orchard.

The hyperparameters for "EfficientNet+PLIA+SFI" that achieved the best precision are: model (B0), image size (224x224), batch size (8), dropout threshold (0.2), learning

DOI: 10.36838/IJHSR819.2

rate (0.00001), and seed value (80). In contrast, the hyperparameters for YOLOv11 that achieved the best precision are: model ("yolov11n-cls"), image size (288), batch size (24), cls value (1.8), dropout threshold (0.1), learning rate (0.0002), and label smoothing value (0.2), respectively.

As shown in Table 1, for both models, the precision is well above 70%, indicating that the models are effective at determining the presence of fully hidden fruits, thereby addressing Question 1. The F1 scores for them are around 0.9, showing a strong balance between precision and recall.

Table 1: Comparison of Classification Results.

Method	Precision	Recall	F1
EfficientNet + pLIA + SFI	0.8636	1.0000	0.9157
YOLOv11 + pLIA + SFI	0.9189	0.8718	0.8947

As an example, the training results of EfficientNet are shown in Figure 9. The plot on the top shows that the training and validation accuracies converge over epochs, while the figure on the bottom shows the results of the confusion matrix of the validation.

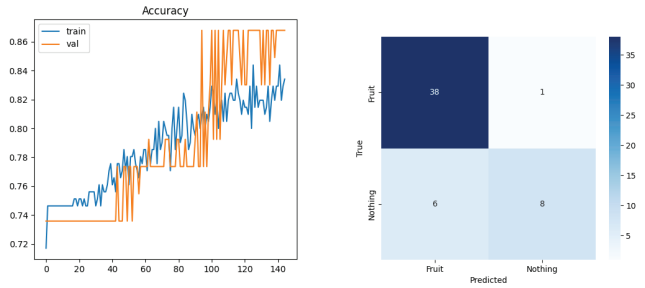


Figure 9: An example training output from EfficientNet. On top is the graph of model accuracy per epoch. At the bottom is the confusion matrix associated with the dataset that the model was trained on.

2. Object detection results:

Three models, YOLOv11, YOLOv12, and RF-DETR, empowered by PLIA and SFI, are trained to detect fully occluded fruits in the mock tomato orchard.

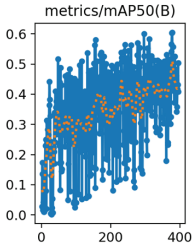
The best results achieved using these three models are shown in Table 2. The hyperparameters used to achieve the best YOLOv11 model are a learning rate of 0.0002, image size of 480, classification loss of 1.8, batch size of 16, dropout threshold of 0.1, and label smoothing of 0.2; whereas YOLOv12 utilizes a learning rate of 0.0001, image size of 480, classification loss of 1.8, batch size of 16, dropout threshold of 0.1, and label smoothing of 0.2. The optimal hyperparameters for RF-DETR are a learning rate of 0.0002, a backbone learning rate of 0.0000001, a batch size of 4, gradient accumulation steps of 4, an image size of 336, a weight decay of 0.0006, and 100 queries.

As shown in Table 2, the mAP50 scores of all models are greater than 65%, which means that the models are effective in determining the number and location of fully hidden fruits, addressing Question 2. For YOLOv11 and YOLOv12 cases, their precision numbers are also higher than 70% in this very challenging task, addressing Question 1.

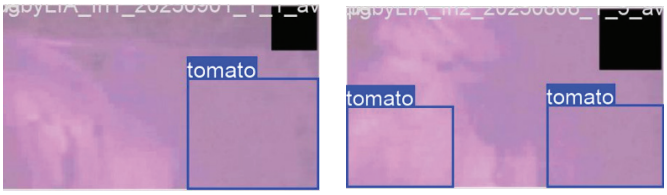
Table 2: Comparison of Object Detection Results.

Method	Precision	Recall	mAP50
YOLOv11 + pLIA + SFI	0.7421	1.0000	0.6874
YOLOv12 + pLIA + SFI	0.7683	0.9870	0.6583
RF-DETR + pLIA + SFI	N/A	N/A	0.6620

Figure 10 shows several plots generated from the training run of YOLOv12 with one hyperparameter setting. The mAP50 value in part (a) oscillates while gradually increasing, and the model parameters corresponding to the best mAP50 value in these epochs are saved. Parts (b), (c), and (d) show three bounding box examples in which tomato locations are successfully identified. In part (e), there is only one tomato; however, YOLOv12 produced two false positives.

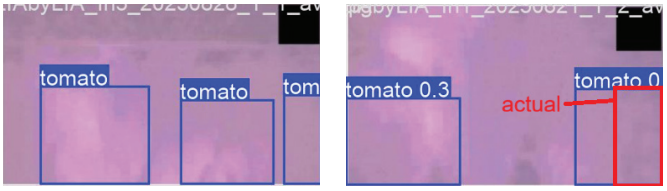


a) The mAP50 value per epoch.



b) One fully hidden fruit with a bounding box.

c) Two fully hidden fruits with two bounding boxes.



d) Three fully hidden fruits with three bounding boxes.

e) There is only one fully hidden fruit; however, the model fails to find it; instead, two false positives are identified.

Figure 10: An example of training output from the YOLOv12 detection model.

3. Model size:

One of the hyperparameters being tuned is the image size for each model. As shown in Table 3, the image size needed for classification is much smaller than that of object detection. Furthermore, the sizes of the trained models are in general very small, ranging from 5.23MB to 19.9M, except the trained RF-DETR model, of which the size is 352MB. This reflects the deployability of each model on small devices. For example, a model such as YOLOv11 for object detection, with a size of only 5.23 MB, can be more easily deployed than a model like RF-DETR, which has a size of 352 MB. This is due to hardware constraints, as a model like RF-DETR may be too computationally heavy to run on a basic Raspberry Pi™.

4. Ablation tests:

Four ablation tests are conducted, aiming to illustrate the contributions of PLIA and SFI to the key performance of the two classification and two object detection models. In each ablation test, two hyperparameter settings are shown: the best result reported previously and a selected (non-best) result. Here, "baseline" refers to training an AI model using raw im-

ages without PLIA and SFI. "PLIA" denotes the case where the raw images are processed using the PLIA method, while "PLIA+SFI" refers to the case when both the PLIA and SFI video/image pre-processing methods are applied.

Table 3: Comparison of Model Complexity.

	Classification		Object Detection		
	YOLOv11	EfficientNet	YOLOv11	YOLOv12	RF-DETR
Image Size (pixels)	288	320	480	480	336
Model Size (MB)	3.04	15.8	5.25	5.25	352

It is worth mentioning that "precision" and "mAP50" are the primary performance metrics focused on during the training for classification and object detection, respectively.

The ablation test result for classification using YOLOv11 is listed in Table 4. The first three rows compare the performance improvement achieved using the proposed video/image pre-processing methods with the hyperparameters that yield the best precision, as discussed previously. "PLIA" outperforms the "baseline" across all three metrics, and "PLIA+SFI" further improves precision and F1 score. These indicate that the proposed methods effectively enhance the models' ability to make more accurate predictions. The last three rows in Table 4 show the ablation test for one selected hyperparameter setting, in which learning rate, image size, batch size, drop rate, and label smoothing are set to 0.0002, 640, 8, 0.1, and 0.2, respectively. As shown, the "PLIA+SFI" strategy improves the precision and F1 score, whereas "PLIA" only achieves the same precision and F1 score as "baseline."

The EfficientNet ablation test is shown in Table 5. The first set of three rows represents the best results for the "baseline", "PLIA", and "PLIA+SFI" cases. These show that the model improves the precision value from the "baseline" case to the "PLIA" case, and from the "PLIA" case to the "PLIA+SFI" case. The second set of three rows represents the results achieved with a set list of hyperparameters, which are a learning rate of 0.0002, an image size of 320, and a batch size of 8. This shows that the addition of "PLIA" and "PLIA+SFI" improves both precision and F1 score.

Table 4: Ablation Test 1 – Classification (YOLOv11).

	Precision	% Increase	Recall	% Increase	F1	% Increase
Baseline*	0.8500	N/A	0.8718	N/A	0.8608	N/A
pLIA*	0.8750	2.9%	0.8974	2.9%	0.8861	2.9%
pLIA+SFI*	0.9189	5.0%	0.8718	-2.9%	0.8947	1.0%
Baseline**	0.7358	N/A	1.0000	N/A	0.8478	N/A
pLIA**	0.7358	0.0%	1.0000	0.0%	0.8478	0.0%
pLIA+SFI**	0.7500	1.9%	1.0000	0.0%	0.8571	1.1%

(*: The best result) (**: Another selected result)

The ablation tests for object detection are shown in Tables 6 and 7. First, the performance comparison for the YOLOv11 object detection mode is displayed in Table 6. The best results for "baseline", "PLIA", and "PLIA+SFI" with their respective hyperparameters listed previously are shown in the first three rows. From these, it can be observed that "PLIA" improves from "baseline" and "PLIA+SFI" improves from "PLIA" across all three metrics. Therefore, the proposed methods are effective at enhancing the model's performance. The last three rows in Table 6 were trained with the same hyperparameters to conduct an ablation test. The hyperparameters are learning rate 0.0002, image size 320, classification loss 1.8, batch size 8, dropout 0.1, and label smoothing 0.2, respectively. As shown here, the "PLIA+SFI" strategy yielded an improvement for precision and mAP50, while the method with only "PLIA" yielded an improvement on mAP50 but a decline in precision from "baseline".

Second, a similar ablation test was conducted for the YOLOv12 model, as shown in Table 7. Rows 1-3 show the best results for the "baseline", "PLIA", and "PLIA+SFI" cases, each with its own distinct set of hyperparameters as shown previously. These results show that each addition of the proposed method improves the mAP50 of the model. Rows 4-6 show the performance of the "baseline", "PLIA", and "PLIA+SFI" methods on the same selected set of hyperparameters, with 0.0002 (learning rate), 480 (image size), 1.8 (classification loss), 24 (batch size), 0.1 (dropout threshold), and 0.2 (label smoothing), respectively. The results unveil that the proposed method enhances the precision, recall, and mAP50.

Table 5: Ablation Test 2 – Classification (EfficientNet).

	Precision	% Increase	Recall	% Increase	F1	% Increase
Baseline*	0.8478	N/A	1.0000	N/A	0.9176	N/A
pLIA*	0.8485	0.1%	1.0000	0.0%	0.9127	-5.3%
pLIA+SFI*	0.8636	1.8%	1.0000	0.0%	0.9157	3.3%
Baseline**	0.8125	N/A	1.0000	N/A	0.8966	N/A
pLIA**	0.8444	3.9%	0.9828	-1.7%	0.9084	1.3%
pLIA+SFI**	0.8636	2.3%	0.9744	-0.9%	0.9157	0.8%

(*: The best result) (**: Another selected result)

Table 6: Ablation Test 3 – Object Detection (YOLOv11).

	Precision	% Increase	Recall	% Increase	F1	% Increase
Baseline*	0.6182	N/A	1.0000	N/A	0.6190	N/A
pLIA*	0.7179	16.1%	1.0000	0.0%	0.6648	7.4%
pLIA+SFI*	0.7421	3.4%	1.0000	0.0%	0.6874	3.4%
Baseline**	0.8167	N/A	1.0000	N/A	0.4807	N/A
pLIA**	0.6761	-14.1%	1.0000	0.0%	0.5358	5.5%
pLIA+SFI**	0.8681	19.2%	1.0000	0.0%	0.6112	7.5%

(*: The best result) (**: Another selected result)

Table 7: Ablation Test 4 – Object Detection (YOLOv12).

	Precision	% Increase	Recall	% Increase	F1	% Increase
Baseline*	0.6987	N/A	1.0000	N/A	0.5910	N/A
pLIA*	0.8384	20.0%	0.9091	-9.1%	0.6224	5.3%
pLIA+SFI*	0.7683	-8.4%	0.9870	8.6%	0.6583	5.8%
Baseline**	0.6279	N/A	1.0000	N/A	0.5460	N/A
pLIA**	0.7299	16.2%	1.0000	0.0%	0.5888	7.8%
pLIA+SFI**	0.8787	20.4%	1.0000	0.0%	0.6046	2.7%

(*: The best result) (**: Another selected result)

In summary, the proposed PLIA and SFI strategies improve the precision and mAP50 values in classification and object detection tasks, respectively.

Limitations

• The current mock tomato orchard only has 1 or 2 layers of leaves covering occluded fruits. In real farm conditions, leaf covers can have more than 2 leaves overlapped on top of a fruit. The used 940nm NIR diodes, which only have a 70mW maximum power, may not penetrate thick leaf covers, and the reflected photons may also scatter. Therefore, in real scenarios, a more powerful NIR diode or laser, or even a millimeter radar, can be utilized. However, the image/video pre-processing methods and the AI models should still be effective.

• The computational cost of the LPF used in PLIA is high because it must iterate through every pixel of an image twice. The current PLIA software is implemented in MATLAB, and in certain scenarios, the execution time of the applied LPF can extend to several minutes. A potential solution is to adopt a computationally efficient LPF and mitigate any resulting performance degradation in the AI models by training on a larger image dataset.

• The performance of the current AI models, while exceeding the thresholds, has not reached 0.95 in precision or mAP50. The simplest way to address this issue is to increase the size

DOI: 10.36838/IJHSR819.2

of the image dataset. Alternatively, new SFI methods can be developed to improve model precision. An instance of this involves predicting the average RGB values of scenarios without fully hidden fruits, which are treated as baseline numbers. The SFI method is then to compute the difference between the average RGB value and that baseline.

Conclusion

In this study, a combination of NIR, image/video pre-processing approaches, and deep AI learning models is proposed to tackle a very challenging problem: detecting and classifying tomatoes that are fully hidden behind leaves. Unlike many other object detection problems, NIR signals reflected from fully occluded fruits are extremely weak and embedded in a large magnitude of noise. To enhance the performance of AI models, a PLIA is developed to extract weak NIR signals from noise and produce an image. After that, the average RGB value of the generated image is used as the side length of a black square, which is injected into the image. This step provides the subsequent AI models with an additional visual cue indicating whether tomatoes are fully hidden behind leaves. Seventy-seven experiment videos obtained in a custom-built, mock tomato orchard are used in training the selected CNN and vision transformer models. All these AI models are able to tell whether there are fully occluded fruits more than 86% of the time. The mAP50 values for object detection are above 0.65. To the best of the authors' knowledge, this approach is low-cost and effective in determining the presence of fully hidden fruits. The ablation tests demonstrate the effectiveness of the proposal PLIA and SFI pre-processing methods in improving the performance of the AI models for classification and object detection.

Acknowledgments

The author would like to thank Dr. Shawn Putnam for the introduction of the lock-in amplifier and coupled image analysis concept. The author would also like to thank Dr. Chen Chen for his thorough proofreading and editing of the manuscript.

References

1. Zhai, Z., Martínez, J., Beltran, V., Martínez, N., Decision support systems for agriculture 4.0: survey and challenges. *Computers and Electronics in Agriculture*, **2020**, *170*, 105256. DOI: 10.1016/j.compag.2020.105256.
2. Chandrakala, P., Badholia, R., Arunkumar, E., Victor, M., Maheswaran, M., Choubey, A., A design of agricultural robotics for the use of sowing and planting. *Proceedings of the 4th International Conference on Advanced Computing, Innovation Technology and Engineering (ICACITE)*, **2024**, 345–349. DOI: 10.1109/ICACITE60783.2024.10616892.
3. Esfandiari, Y., Nagasubramanian, K., Fotouhi, F., Schnable, P., Ganapathysubramanian, B., Sarkar, S., Distributed deep learning for persistent monitoring of agricultural fields. *Neural Information Processing Systems (NeurIPS 2021 AI for Science Workshop)*, Virtual, **2021**. https://openreview.net/forum?id=rctFXFsFvbI.
4. Sharma, N. A. P., Dadheech, P., Aneja, S., Predicting agriculture yields based on machine learning using regression and deep learning. *IEEE Access*, **2023**, *11*, 111255–111264. DOI: 10.1109/ACCESS.2023.3321861.
5. Barburiceanu, B. O. R. L. S., Meza, S., Terebes, R., Convolutional neural networks for texture feature extraction: applications to leaf disease classification in precision agriculture. *IEEE Access*, **2021**, *9*, 160085–160103. DOI: 10.1109/ACCESS.2021.3131002.
6. Vougioukas, S., Agricultural robotics, *Annual Review of Control, Robotics, and Autonomous Systems*, **2019**, *2*(1), 365–392. DOI: 10.1146/annurev-control-053018-023617.
7. Fontani, M., Luglio, S. M., Gagliardi, L., Peruzzi, A., Frasconi, C., Raffaelli, M., and Fontanelli, M., A systematic review of 59 field robots for agricultural tasks: applications, trends, and future directions, *Agronomy*, **2025**, *15*(1), 2185. DOI: 10.3390/agronomy15092185.
8. Houetohossou, S., Houndji, V., Hounmenou, C., Sikirou, R., Kakaï, R., Deep learning methods for biotic and abiotic stresses detection in fruits and vegetables: state of the art and perspectives. In *Artificial Intelligence in Agriculture*, **2023**. *9*, 46-60. DOI: 10.1016/j.aiia.2023.08.001.
9. Moldvai, L., Nyéki, A., Innovative computer vision methods for tomato (Solanum lycopersicum) detection and cultivation: a review. *Discover Applied Science*, **2025**, *7*(975). DOI: 10.1007/s42452-025-07613-x.
10. Cho, J. W., Tin, T. T., Tu, L. H. T., Park, G. S., Chien, V. D., Kim, J. W., A nutrient deficiency prediction method using deep learning on development of tomato fruits. in *Proceedings of IEEE Conference*, **2018**, 338–341. DOI: 10.1109/iFUZZY.2018.8751688.
11. Koirala, A., Walsh, K., Wang, Z., McCarthy, C., Deep learning method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, **2019**, *162*, 219-234. DOI: 10.1016/j.compag.2019.04.017.
12. Xiao, F., Wang, H., Xu, Y., Zhang, R., Fruit detection and recognition based on deep learning for automatic harvesting: an overview and review. *Agronomy*, **2023**, *13*(6), 1625. DOI: 10.3390/agronomy13061625.
13. Guo, J., Yang, Y., Lin, X., Memon, M. S., Liu, W., Zhang, M., Sun, E., Revolutionizing agriculture: real-time ripe tomato detection with the enhanced Tomato-YOLOv7 system. *IEEE Access,* **2023**, *11*, 133086–133098. DOI: 10.1109/ACCESS.2023.3336562.
14. Duong, L., Nguyen, P., Sipio, C., Ruscio, D., Automated fruit recognition using EfficientNet and MixNet. *Computers and Electronics in Agriculture*, **2020**, *171*, 105326. DOI: 10.1016/j.compag.2020.105326.
15. Tan, M., Le, Q. V., EfficientNetV2: smaller models and faster training, arXiv preprint, 2021. https://arxiv.org/abs/2104.00298.
16. Lawal, M., Tomato detection based on modified YOLOv3 framework. *Scientific Reports*, **2021**, *11*, 1447. DOI: 10.1038/s41598-021-81216-5.
17. Yang, S., Wang, W., Gao, S., Deng, Z., Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin transformer. *Computers and Electronics in Agriculture*, **2023**, *215*, 108360. DOI: 10.1016/j.compag.2023.108360.
18. Xiao, B., Nguyen, M., Yan, W., Fruit ripeness identification using transformers. *Applied Intelligence*, **2023**, *53*, 22488–22499. DOI: 10.1007/s10489-023-04799-8.
19. Sa, I., Lim, J., Ahn, H., MacDonald, B., DeepNIR: datasets for generating synthetic NIR images and improved fruit detection system using deep learning techniques, *Sensors*, **2022**, *22*(13), 4721. DOI: doi.org/10.3390/s22134721.
20. Jocher, G., Qiu, J., Chaurasia, A., *Ultralytics YOLO*, Software, 2023. https://github.com/ultralytics/ultralytics.
21. Robinson, I., Robicheaux, P., Popov, M., *RF-DETR*, Software, 2025. https://github.com/roboflow/rf-detr

22. *PerkinElmerTM Instruments, Low level optical detection using lock-in amplifier techniques*
23. Xu, A., An integrated lock-in amplifier and near infrared method for finding fruits fully behind leaves in a homemade testbed, *International Journal of High School Research*, **2026**, *6*(2), 1–6. DOI: 10.36838/v8i2.1.
24. Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T., Cubuk, E., Le, Q. V., Zoph, B., Simple copy-paste is a strong data augmentation method for instance segmentation, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, **2021**, 2917–2927. https://openaccess.thecvf.com/content/CVPR2021/html/Ghiasi_Simple_Copy-Paste_Is_a_Strong_Data_Augmentation_Method_for_Instance_CVPR_2021_paper.html?ref=https://githubhelp.com
25. Gené-Mola, J., Gregorio, E., Cheein, F. A., Guevara, J., Llorens, J., Sanz-Cortiella, R., Escolà, A., Rosell-Polo, J., Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced air flow, *Computers and Electronics in Agriculture*, **2020**, *168*, 105121. DOI: doi.org/10.1016/j.compag.2019.105121.
26. Couture, A., Adams, C., Fonseca, R., Schubert, J., Mastronardi, R., Backscatter imaging for precision agriculture, *US Patent US10712293B2*, **2020**.
27. Mahmud, M., Zahid, A., He, L., Choi, D., Krawczyk, G., Zhu, H., LiDAR-sensed tree canopy correction in uneven terrain conditions using a sensor fusion approach for precision sprayers. *Computers and Electronics in Agriculture*, **2021**, *191*, 106565. DOI: 10.1016/j.compag.2021.106565.
28. Shiraz, Z., Khan, U., Shahzad, M., FruitSight: a millimeter wave radar based approach to detect occluded fruits. *Proceedings of IEEE Conference*, **2024**, 270–278. DOI: 10.1109/MASS62177.2024.00044
29. Kim, S., Kim, D., Fruit quantity measurement system and method. *KR Patent 12-541969*, **2020**.
30. Parr, B., Legg, M., Bradley, S., Alam, F., Occluded grape cluster detection and vine canopy visualization using an ultrasonic phased array, *Sensors*, **2021**, *21*(6), 2182. DOI: 10.3390/s21062182
31. Weisstein, E. W., Fourier series square wave, *Wolfram MathWorld*. https://mathworld.wolfram.com/FourierSeriesSquareWave.html (accessed April 13, 2026).
32. Basel problem, *Wikipedia*. https://en.wikipedia.org/wiki/Basel_problem (accessed April 13, 2026).

■ Author

Aiden Xu is a high school student at Winter Springs High School, Winter Springs, FL. He is interested in conducting different types of research during his spare time and has been participating in the Science and Engineering Fair since the 6th grade.

DOI: 10.36838/IJHSR819.2

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Extracting Influence: The Australian Mining Industry and National Climate Policy

Andy Guan

Redlands School, 272 Military Road, Cremorne, New South Wales, 2090, Australia; agcomp09@icloud.com
Mentor: Mijoo Shin

ABSTRACT: The growing urgency of climate action to avert the disastrous consequences of climate change has influenced the Australian Government to announce its commitment to fighting climate change, exemplified by its net-zero emissions by 2050 commitment and the signing of the Paris Agreement. However, despite the environmental impacts of mining, the economic and political influence of the mining industry in Australian politics has created political stagnation and climate inaction, undermining Australia's commitment to climate leadership and accountability. While many previous assumptions about climate inaction have pointed towards the economic cost of a decrease in GDP, the current energy production dependence on fossil fuels, the potential effect on national employment levels, or the defunding of climate science programs by previous governments, many omit the political influence of the mining sector in pressuring mining-friendly legislation or approvals. This paper argues that political spending of the Australian mining industry influences Australian climate policy outcomes. The Australian mining industry, through spending such as political donations and advertising campaigns, has led to the lack of climate legislation and continued mining expansion, which undermines the government's promise of climate action.

KEYWORDS: Political Science, Public Policy, Australia, Environmental Politics, Mining Industry, Climate Change.

Introduction

The Australian Government has committed to the legally-binding Paris Agreement, which states that nations commit "to limit global warming to 1.5°C, greenhouse gas emissions must peak before 2025 at the latest and decline 43% by 2030,"[1] and has set out a clear target for net-zero emissions by 2050.[2] Despite its commitments to climate action, which would require a halt of further new gas, coal or oil projects, it has avoided major legislation that reduces or economically harms the mining industry due to successful political lobbying. The continued support of the mining industry and its expansion is one of the reasons why Australia is currently not on track to meet its climate objectives.[3] This contradicts the global consensus that the mining industry is detrimental to the environment and thus active steps should be taken to switch to renewable energy production or at least stop or rescind new approvals for expansion.[4]

Even within Australia, there have been calls for urgent climate action, as many recognize the threat that climate change poses to humanity and the Earth. Many Australian citizens have been affected by climate change, especially through the increasing frequency and severity of natural disasters, such as the 2019 Black Summer Bushfires. Over 70% of Australians said they are concerned about climate change, and a majority of Australians (55%) said that they support Australia's commitment to global objectives of limiting global temperature rise to 1.5°C, including the requirement to terminate new mining projects to meet those goals. Many have also recognized the adverse effects of mining on the environment, with 56% of Australians being actively aware that opening new gas, coal, or oil projects will make it more difficult for Australia to reduce greenhouse gas emissions by 43% by 2030. Similarly, many also agree that to better protect the environment and meet its targets, Australia cannot continue with the status quo, and 66% of Australians think that the government should plan to phase out coal mining and transition into other industries.[5]

Recent studies on the causes of climate inaction in Australia include papers discussing the various framings of climate discourse, the role of narratives and conservatives, economic dependence on the industry, ineffective government institutions, and more. However, previous research uses frameworks that are too narrow in scope, such that they do not always reflect the general patterns and causation. Also, some arguments and trends observed no longer reflect reality due to the rapidly changing nature of public opinion and climate discourse. Finally, the existing research discusses non-essential factors, rather than the core reasons. This paper contributes to the existing research by looking at political spending, specifically political donations and media campaign spending by the mining industry on national climate policy outcomes.

This paper aims to explore how political spending by the mining industry has created climate policy inaction based on the theoretical framework that politicians are disproportionately influenced by financiers. It introduces a central causal mechanism and utilizes various historical and modern examples to understand the relationship between political spending and climate policy inaction. Through the analysis of political donations and media spending patterns and their impact on shaping the major parties and political outcomes, I will argue that the political spending of the Australian mining industry is a key factor that has influenced climate policy inaction in Australia.

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.3

In the next section, I will discuss existing literature and its limitations. In the third section, I will explain the theoretical framework of the paper, including the causal mechanism and the mining industry in Australia. In the following section, I will investigate the impact of political donations and media spending on climate policy and the subsequent policy inaction. Finally, I will close with a brief conclusion.

Literature Review

Framing:

Existing studies point to the political framing of climate discourse as the primary cause for policy inaction. These studies assume that political discourse engaged by politicians within policy-making bodies represents the general view of the constituents. In particular, scholars highlight the relationship between economic framing and the failures of both international organizations and domestic policymaking in passing effective regulations. Christoff argues that Australia's policy inaction has been caused by waning public support and awareness because of failures of joint, international commitments for urgent and domestic policy implementations.[6] Christoff recognizes a shift in public recognition of climate change, and a subsequent shift in economic discourse from "old" to "new," signifying an explicit recognition of climate change. However, Christoff asserts that failures of international commitments undermined political momentum for decisive climate policy and discourse. The failure of international commitment, specifically through the examination of the 2009 Copenhagen climate negotiations, has spurred political opposition against climate-positive legislation in favor of the economic argument, which was evident in the divisive debates for carbon pricing. The lack of urgency in climate action suggests that the consensus disregarded the exigency and that politicians acted based on the stances of their constituents. Furthermore, media focus on divisive domestic debates, such as Carbon Pricing, further eroded public support for climate action.

However, since then, there has been a significant shift in both public awareness and urgency as well as public support for tougher regulation, particularly among the younger generations. This means that the majority of Australians no longer disregard climate as a problem. The shift is evident in significant civil participation by environmental activists to influence governmental policy, significant youth protest movements such as Fridays for Future and Extinction Rebellion, as well as other nationwide and grassroots climate justice protest movements that represent a shift in public opinion. This is supported by data that suggests that over 70% of Australians are concerned about climate change.[5] While applicable at the time of writing (2013), significant shifts in public opinion and intensive civil participation for stricter environmental policies mean that such an argument is no longer as accurate.

Role of Conservatives:

Other studies also point to the power of conservatives within Australia's major parties, specifically within the Liberal-National Party Coalition, which has dominated Australian politics in the 2000s and 2010s. Crowley argues that there is division within both the Liberal-National Party Coalition and the Labor Party in the stance of the dominant parties on climate action. Conservatives utilize climate policy differences to incite internal political dissent, resulting in conservatives who are staunchly in support of mining being able to exert influence within the party. This has resulted in a pro-mining-driven narrative, with internal party politics and warring internal factions being more instrumental in dictating positive climate policy. Even with public support, it has not swayed the government from adopting climate policies.[7] While conservatives do exert significant influence, there have been instances where pro-climate policy has been implemented or promised because of public support and opinion. An example is Howard's 2007 electoral commitment to an Emissions Trading Scheme, despite being historically against it and despite incurring significant protest and challenges from the conservatives.[8] Another example is the federal approval of Snowy Hydro 2.0 under the Turnbull government, providing clean energy, despite facing intense resistance from conservatives due to various reasons, such as opposition to the implementation of clean energy and the high projected cost.[9] This demonstrates that while conservatives have the power to delay or hinder policies from being introduced, ultimately, public pressure and opinion have the power to force action. Moreover, it does not explore the impact of external sources of influence on the views of the conservatives, which is then reflected in their position against climate action. In this case, mining lobby groups help sway conservatives to continue their anti-environmental policy despite mounting public pressure.

Ineffective Government Institutions:

Others point to the role of ineffective domestic institutions in supporting the introduction and implementation of effective climate policy in Australia. MacNeil defines these institutions as "formal and informal procedures, routines, norms, and conventions embedded in the organizational structure of the polity or economy" under the federal government. MacNeil argues that Australian climate policies have been largely hindered by the political system locked in a highly polarized debate around anthropogenic climate change and climate policy. This creates a limited number of institutions that lack the agencies and resources to create decisive policy to combat climate change. This results in weak national policy, such as the dismantlement of various institutions under the Abbott government. These institutions are thus unable to protect the national climate policy against the toxic national debate on climate change, which results in varying policies dependent on the current party in power. MacNeil further argues that the climate policies that have been introduced by such institutions are unable to take decisive action as they need to appease both individuals and groups that support climate action, but also anti-climate constituents, especially within elements of the conservative Liberal party.[10]

However, government institutions have not been as ineffective as they seem. They have been quick to adopt climate-related policy, which is evident in the responses to proposed climate legislation and requests for mining expan-

sion. Government institutions have been favorable to the mining industry, as can be seen by the rejection of drastic new environmental laws and approvals of mining expansion by both major parties. This is largely attributed to the political influence wielded by the industry. Even if these legislations are not pro-climate, they are cases where government institutions have acted efficiently, influenced by the industry. Moreover, these arguments also do not address the core reason why such climate discourses experience such high polarization. The polarization has caused division within both the Labor and Liberal-National Coalition Party, with many party members becoming more conservative. The mining industry, through lobby groups, swiftly passed policies related to the expansion of mining.

Methods

Theoretical Framework:

Past studies about campaign contributions argue that people who donate money to political parties view it as an investment for their interests, and once the party that they supported is elected, they will gain something in return.[11] These could be more beneficial policies, an amendment of legislation, a repeal of a law, preventing a proposed law debated in Parliament from being passed, tax cuts, and more. Joo-Cheong Tham argues that "when judging what is in the 'public interest', the minds of politicians will be skewed towards the interests of their financiers."[12] These political donations lead to higher engagement between legislators and lobby groups, leading to the interests of certain political groups being heard more than others, skewing political outcomes in favor of donors. These political donations are important resources for political parties to run campaigns and to fund other party ventures, with parties increasingly relying on them. These donations financially pressure legislators to act in favor of donors to secure future funding, informally binding politicians to the interests of donors. Similarly, other types of political spending, such as public relations campaigns and advertising investments, can shift public opinion in favor of the view of lobbyists and directly pressure governments to act in accordance with the interests of the lobbyists.

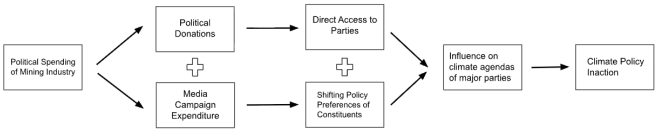


Figure 1: Political Spending of the Mining Industry on National Climate Policy Inaction. It represents the causal mechanism of how political spending by the mining industry has impacted Australian climate policy outcomes.

Figure 1 represents the core argument of my paper, demonstrating the causal mechanism of political spending by the mining industry and its subsequent impact on Australian climate policy. The political spending comprises donations to the major parties and media campaign expenditure. Political donations increase access to the major parties and allow preferential hearing of the interests of the mining industry. This permits them to directly influence the parties' policy stances. At the same time, media campaign expenditure shapes the policy preferences of constituents, which indirectly pressures the government to introduce favorable policies to the mining industry. The combination of political donations and media campaign expenditure influences the climate agendas of the major parties. This has a direct implication on national policy-making, resulting in climate policy inaction by the Australian Government.

Case Study: Mining Industry in Australian climate politics:

The mining sector in Australia grew considerably into large, industrialized corporations that run massive mines. Growth in the industry was accentuated during the energy and mineral demand boom of the 2000s.[13] By 2020, Australia was one of the largest producers of coal and had proven reserves of 150.227 billion tons, a staggering 37.6% of global iron ore reserves, around 30% of global uranium reserves, and had large reserves of many other precious metals and minerals such as opal and nickel. Australia also remains one of the largest natural gas and oil producers globally, with 151.7 billion cubic meters and 140.89 million barrels produced in 2023, respectively.[14]

The profits of these companies grew considerably with increasing global demand due to mass industrialization, especially among developing industrial economies, notably China. Substantial profits helped investment into the industry and supported the growth of large mining giants such as BHP and Rio Tinto.[15] Through consolidation and involvement as a stock company, these corporations have created a very high market capitalization, highlighting their wealth.[16]

Currently, mining makes up around 14% of the National GDP,[17] around 12% of National Output,[18] and is the most exported product. In the 2022-2023 financial year, it reaped a revenue of $241.9 billion, accounting for two-thirds of the national export revenue.[19] In 2025, the mining industry employs around 324,800 individuals,[20] making it an important factor for consideration when approving expansions and has been touted as a key benefit of mining by the industry.

However, mining causes significant environmental degradation both in the immediate vicinity and in its contribution to anthropogenic climate change. It results in considerable mechanical damage to the surrounding landscape and creates acid mine drainage. Mining activities cause substantial soil degradation, leading to acidic soil with lower soil pH,[21] and contaminating local waterways through the leaching of heavy metals and tailings into waterways, with abandoned mines also contributing to water quality degradation.[22] Moreover, considerable amounts of harmful particulates and gases, such as Sulphur dioxide, degrade air quality,[23] and contribute to global emissions through significant amounts of greenhouse gases, such as methane and carbon dioxide, released as a result of mining operations.[24] This results in negative environmental outcomes for the local and global environment.

The large profits of the industry and its impact on Australia's economy allow it to exert large amounts of influence in the political landscape of the country, especially among the more conservative parties such as the Liberal-National Coalition. This has resulted in the development of a close relationship between the government and lobbyists, notably between the

DOI: 10.36838/IJHSR819.3

Liberal National Party Coalition and the industry, with greater government support of extraction and exploration through subsidies.[25] The State and Federal Governments have actively supported the growth and development of this industry by providing substantial financial subsidies and tax breaks. The government has also supported the industry by granting mining tenements that give companies the right to construct and extract from mines.[26] In 2015, the Australian Government provided $4 billion in subsidies, yet that has considerably grown in the following decade to around $14.5 billion in the year 2023-2024 to coal mines, oil, and gas operations.[27]

These financial supports divert funding from other aspects of federal spending, including climate research or response, and make fossil fuels a cheaper option than renewable alternatives, disincentivizing a national shift towards renewable and cleaner energy sources and technologies.[28] These subsidies are also very unpopular with the Australian people, with only 7% of Australians supporting government subsidies for mines according to a survey conducted by the Australian Institute.[5]

Results and Discussion

In Australia's case, political spending by mining corporations aims to exert more influence and pass new legislation that supports the expansion of mining and prevents laws that would economically harm the industry, such as stricter regulations on mining approvals or laws that prevent further expansion. This highlights the relationship between the government and the industry, where industrial lobbyists, who act in the interests of mining companies, shape policy. The vested interests of political spending by mining corporations are economic in nature, striving to increase their profits and benefit investors. Such spending has allowed the industry to exert a significant influence on the government to enact favorable legislation, contributing to climate inaction. This directly contradicts the government's climate promises and the general sentiment within the Australian population that supports stronger governmental regulation and intervention to better protect the environment.

Political Donations:

Mining corporations have regularly used political donations to exercise considerable influence on parties' policy agendas, most notably the conservative parties, such as the pro-mining Liberal-National Coalition. This has been possible because political parties in Australia are primarily run by private funds, especially for federal election campaigns. The impact of financial donations on the parties is that the interests of the mining industry are prioritized. The mining industry uses this money to access business forums and other political events organized by the major parties, influencing the party's policy.[29] Political donations by mining corporations such as Rio Tinto, Glencore, and BHP have allowed unparalleled access to the political and economic events of the major parties.[30]

Financial donations have helped secure meetings between the mining lobby and legislators and bureaucrats, where donors aim to persuade politicians to introduce or block policy in favor of the mining industry. This allows the constant engagement of politicians with issues raised by donors through frequent interactions with politicians and mining lobby groups.[31] As a consequence, it causes preferential hearing of the interests of political donors over other concerns raised by non-financiers, such as environmental activists, despite public support of environmental action. An official Senate inquiry into the corrupting effects of political donations highlights the issue that Australians have recognized as a threat to proper and fair democracy and its effect on skewed policies that favor the interests of the mining industry.[12]

The most influential lobby group is the Minerals Council of Australia, which pushes for financial subsidies and approvals for the expansion of mining. In an Australian Senate Inquiry into political donations in 2018, the Minerals Council of Australia conceded that it has made political donations to gain access to members of Parliament and directly lobby their interests.[12]

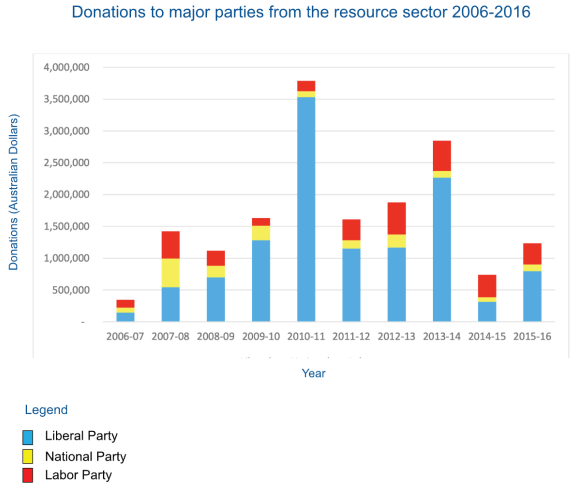


Figure 2: Political donations from the mining industry to major parties (2006–2016). Adapted with permission from reference 32. Copyright 2017 The Australia Institute. Between 2006 and 2016, the majority of political donations from the resource (mining) sector were given to the conservative Liberal-National Coalition, notably during federal election years.

Australia is a multi-party system, but it operates like a system with two major parties, the Australian Labor Party and the Liberal-National Coalition. At each election, one of the two major parties would win a majority of seats in Parliament to form a government. They do not need to compromise with minor parties that are against mining activity, such as the Greens, or with independent candidates. This is why political donations are focused on the Liberal-National Coalition and the Australian Labor Party, because it allows the maximum impact of political donations.

Based on the 2017 discussion paper, *The tip of the iceberg: Political donations from the mining industry* by the Australian Institute (Figure 2), between 2006-2016, the mining industry donated $16.6 million to the 3 major parties. This included 71% of donations made to the Liberal Party, with the remainder divided between the Labor and the National Party. However, current reporting laws allow the mining industry to conceal political donations below a certain threshold and make multiple donations of small amounts without being disclosed.[32] By

making large amounts of donations to the Liberal-National Coalition, the mining industry has influenced the stance of the Coalition. The Liberal Party has constantly supported natural resource exploitation on the grounds of economic growth and energy security. In addition, the mining industry utilizes financial donations as a way to influence the Australian Federal Election outcome. This money is used for campaigning and to sway the climate agenda of the incoming government.

According to Figure 2, there has been a spike in donations in the federal election years: 2007, 2010, and 2013. This observation is consistent with findings of Tello, Hazelton, and Leong, who noted that corporate political donations drop significantly during non-federal election years.[33] The large amount of donations in the 2010 election highlights the motivation of the mining industry to sway political outcomes on legislation that would harm them, especially the mining profit supertax proposed by Labor candidate Kevin Rudd. The exact amount of political donations is likely to be much higher than the declared amount, because donations under $16900 are not required to be disclosed.[34] In 2022-2023 alone, 42% of private income and 22% and 27% of total income of the Liberal and Labor Party, respectively, came from undisclosed sources.[35]

In Figure 2, the skewed distribution of donations towards the Liberal-National Coalition demonstrates the mining industry's attempt to influence the federal election through greater funding to the conservative party. Conservatives are generally more supportive of the mining industry because of its emphasis on the economic benefits of mining. They are more concerned with maintaining the status quo, rather than environmental sustainability, which does not have an immediate effect on the livelihoods of voters. There are greater concerns over the disruption of the mining industry, which has economic implications. By passing laws favorable to the mining industry, the conservatives can also win the support of constituents who are in line with their party's ideology.[36]

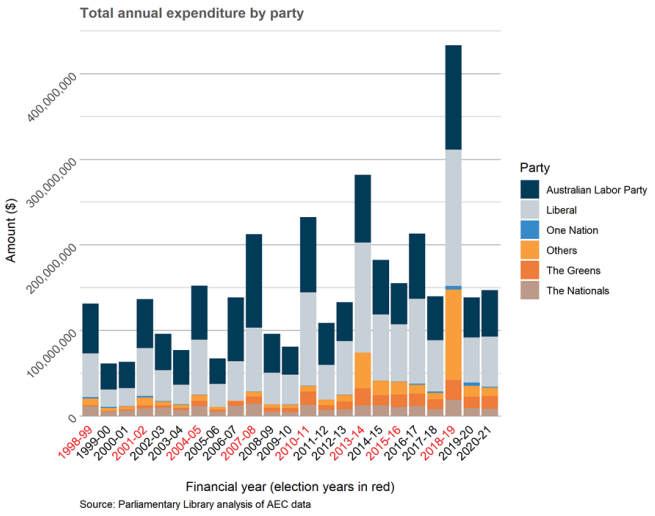


Figure 3: Total annual expenditure by major parties. Reproduced from reference 37 under a CC BY-NC-ND 3.0 AU license. Between 1998 and 2021, the annual political expenditure of major parties increased, particularly in Federal Election years.

The impact of political donations is intensified by the recent phenomenon in which election campaigns are becoming increasingly expensive. Based on the annual expenditure by the major parties between 1998 and 2021 (Figure 3), there is a massive increase in party expenditure during election years. While it is understood that costs increase during election years to fund election campaigns, in recent years, the amount has grown considerably. In addition, the parties need to diversify campaign strategies to attract new types of voters. They now have to launch social media campaigns in addition to using traditional media such as radio and television. Thus, many parties are increasingly reliant on big donations. Those with greater campaign funding can increase their visibility to national audiences, retain supporters, and attract new voters. Budget increases allow active public relations campaigns targeting marginal and swing seats, where competition is very tight and can be more easily won compared to less marginal or safe seats.

Election Year	Total Votes		
	Labor Party	Liberal-National Coalition	Other candidates
1975	36	91	0
1990	78	69	1
2010	72	73	5
2022	77	58	16
2025	94	43	13

Figure 4: Australian federal election results - House of Representatives, 1975–2025. Data adapted from reference 38. Since 1975, seats in the House of Representatives have shifted away from being wholly divided between the Australian Labor Party and the Liberal-National Coalition to a more plural composition, with increased representation of minor parties and independent candidates.

In recent years, the House of Representatives in federal elections has become more plural due to the rise of independents and minor parties. The Greens Party and non-party-aligned Independent candidates are against mining. The Mining Industry has tried to limit its influence and ensure that the Liberal-National Party wins most seats. Figure 4 demonstrates the shifting voting patterns for the House of Representatives from 1975 to 2025, particularly in the years 1975, 2022, and 2025. Figure 4 illustrates the bipartisan voting pattern in 1975 for the House of Representatives, with the vote wholly split between the Australian Labor Party and the Liberal-National Coalition. The bipartisan voter trend allows the parliament to be split between the two parties, making attaining a majority to form a government easier and less complex. However, since the election of the first Independent member to North Sydney, Ted Mack, in 1990, there has been a rise in Independent and Minor party candidates. The major parties declined, and the rise of these other candidates has increased competition. Instead of major parties competing against one another, they now contend against smaller parties and Independents, splitting votes more.

In the 2025 election, based on YouGov, there was a rise in the influence of candidates not representing the Labor Party or Liberal-National Coalition, with an increased projected vote share compared to previous elections.[39] This has created tighter races in all electorates for both major parties. In the 2025 election, non-major party candidates won thirteen seats,

DOI: 10.36838/IJHSR819.3

creating a large crossbench of non-major party parliamentarians,[40] with a dramatic shift from zero seats in the 1975 election to thirteen in 2025. The 2022 election results also reflect this trend, where non-major party candidates won sixteen seats. This is likely to continue in future elections, making more seats unavailable to major parties to obtain a majority to form a government.

This pattern has significantly magnified the importance of campaign spending. The rise of the non-major parties and Independents has meant that major parties must compete with more parties for voters. In the 2025 election of the Sydney electorate, Bradfield, the voter margins were one of the lowest recorded in history, with the Independent candidate winning by 26 votes.

In sum, mining companies have used financial means to influence key parties and their stances on climate policies. They have utilized political donations to support the party most aligned with their interests.

Media Framing:

As climate change has become a more pressing issue, public opinion has shifted towards greater support for environmentally beneficial reforms and a transition towards renewable energy. Public support for government action to address climate change has been a threat to mining corporations. In a report by the Commonwealth Scientific and Industrial Research Organization (CSIRO),[41] 64% of Australians agreed, or strongly agreed, that collective efforts can reduce the negative effects of climate change. Furthermore, 81% were at least somewhat interested in the energy transition. The mining corporations have tried to sway public opinion through advertisements, public relations campaigns, media partnerships, and social media campaigns. Given that in a democracy, the government acts in the interests of voters, shifting public opinion will directly influence policy outcomes.

The mining industry has conducted advertisements that emphasize its role in economic growth and job creation. Advertisements create a perception of minimal effects on environmental impacts.[42] For instance, Woodside Energy, one of Australia's largest mining corporations, framed mining as a way to increase jobs and economically benefit communities. Subsequently, by shifting constituent policy stances, the industry has influenced the climate agendas of the major parties. It successfully brands itself as a company that engages in carbon-reducing behavior and emphasizes continued commitment to a sustainable future by setting out clear net-zero emissions by 2050. It has continued campaigning and seeking approval for mining expansions. It gained approval in May 2025 to extend the operation of its North West Shelf Gas Plant in Western Australia to 2070, 20 years after its net-zero claims.[43] These advertisements attempt to address constituent concerns by assuring people that mining activities are adapting to net-zero goals and are transitioning towards clean mining practices.

In 2025, the Ad Standards Community Panel, an Australian advertisement regulator, found that Hancock Prospecting, a mining conglomerate, had intentionally misled consumers with its unsubstantiated claim that its gas was clean, violating environmental codes that advertisers are required to abide by.[44] The use of media campaigning permits the industry to reframe the environmental consequences of mining and create a perception that mining practices are harmless.

Another instance was the mining industry's threats towards a proposed windfall profit tax, in the context of its increased revenue from higher gas and coal prices following Russia's invasion of Ukraine. Supporters of the windfall tax, such as Richard Denniss argue that "the fact that companies are making these record profits while Australians struggle with the cost of living crisis is likely why almost 9/10 Australians support Government intervention into the energy market, according to…certified research."[45] As a response, the Mineral Council of Australia pushed back, arguing that "we can't afford to see jobs go in regional Australia…we can't afford to see investment slow down." The spokesperson for the Minerals Council of Australia, Tania Constable, announced the intention of the council to start a substantial advertising campaign focused on the economic implications, similar to the one that ended the Mining Super Profits Tax proposed by the Rudd Government, if such a policy were to pass. After the statement was made, the proposal was not passed by the government after careful consideration and consultation with the lobbyists.[46]

Policy Outcome: Climate Inaction:

The mining industry has continued to employ its financial means to block any legislation that would financially affect it or hinder its ability to obtain policy approvals. This often clashes with public opinion, which wants the Australian government to actively take steps to protect the environment and mitigate any further pollution and degradation caused by the mining industry.

Mining lobbyists have exerted significant influence on climate discourse and national policy to combat climate change since the earliest attempts and pledges for environmental action. The first policy promise was during the Labor government under Bob Hawke, which agreed to international cooperation at the 1988 Toronto Conference on Climate Change, pledging a 20% reduction in greenhouse gases from the 1988 levels by 2005. It was followed by the proposal to the cabinet in 1989 and the subsequent failure to adopt it in 1990. Despite the promise of change, policy inaction continued throughout the remaining term of Hawke, the following government under Paul Keating, and particularly during the term of the Liberal Party under John Howard until 2007. Inaction under Keating was largely attributed to political challenges in the wake of the 1990s recession.[47] Under the Howard government, climate policy was under significant influence of mining and energy sector lobbyists, despite pressures from scientists, environmental organizations, and business groups such as the Business Roundtable on Climate Change.[44] Pearse, in his book *High and Dry,* highlights the pervasive nature of representatives of lobby groups of major polluting industries, especially mining, which was referred to as the 'greenhouse mafia', in industry and governmental positions that dictated policy under Howard. The political influence of the mining industry during the Howard

government between 1996 and 2007 prevented climate action that could negatively impact the industry. As a result, there was a decade of significant inaction during a critical period, and the Howard Government attempted to jeopardize international cooperation in climate change. It rejected ratification of international climate agreements such as the 1997 Kyoto Protocols and promised the public that no further goals were being considered.[48] This was despite extensive advocacy from various environmental groups, members of the public, many government personnel, and politicians.[49]

One of the first instances of direct political intervention by mining companies is their response to the proposed super tax on mining profits by Kevin Rudd. While this was not directly related to the environmental issue, it demonstrated the power of the mining industry. In May 2010, Prime Minister Kevin Rudd introduced a highly debated 40% Resource Super Profits Tax on mining companies. Government ministers supported this proposal as it would bring greater government revenue and distribute the benefits of the recent mining boom to citizens. It was further supported by scholars as the key recommendation of the 2010 *Australia's Future Tax System Review Final Report* commissioned by Treasury Secretary Ken Henry.[50]

The mining industry was alarmed by the prospect of a 40% tax on its revenue. Spearheaded by the Minerals Council of Australia and mega-mining corporations such as Rio Tinto and BHP, the mining industry began a costly media campaign amounting to $22.2 million, attacking the tax with contested claims. The media campaign pressured the Labor Party to replace Kevin Rudd with Julia Gillard, who negotiated a much weaker tax on mining profits.[51] In their study, Bell and Hindmoor concluded that the super tax ultimately failed because the mining industry was able to persuade a large number of voters that the tax would be economically deleterious.[52]

The mining supertax would have generated a significant taxation income, amounting to more than $200 billion in the following decade, for the government to spend on other initiatives. This included potential spending on environmental protection and for climate policy implementation, with projects that would benefit Australians, especially because most of the shares and profits of mining companies are owned by foreign investors. While some argued that the high amount of tax is unsustainable, the mining industry reaped immense profits due to the global boom, amounting to $51 billion in 2009-2010 alone, and a continued projected growth for the coming decade.[53] The high profits of the industry would have allowed the industry to afford the tax while maintaining high profit margins.

Moreover, the political influence of lobby groups has marginalized indigenous groups, who opposed mining activities on sacred lands. Indigenous culture is deeply connected to the land, which makes up a fundamental part of their spirituality. They believe in their responsibility in protecting the land as traditional custodians.[54] Indigenous groups campaigned against the construction of the Carmichael Mine, which was proposed by Adani Enterprises, an Indian industrial conglomerate, to the Queensland Government in 2010. Adani submitted an application to construct one of the world's largest greenfield coal mines, with supporting infrastructure in the Galilee Basin on traditional Wangan and Jagalingou lands. From the start, the local Indigenous community vehemently protested the construction over major environmental concerns to their traditional homelands, affecting local groundwater, ecosystems, and endemic species, especially with Adani's history of development that incurred damages to the environment. Local groups rejected Adani's proposed Indigenous Land Use Agreements that would extinguish native title rights. The Indigenous community launched a well-publicized "Adani, No Means No" campaign and took legal action that attracted significant public support. This created mass controversy during the 2017 Queensland State Election, which resulted in a shift in the government's position by removing $1 billion in federal funding to reduce electoral backlash. However, the government continued to support the approval.[55] The mining lobby supported the construction of the mine and sought governmental backing of the project. They secured 191 contacts between them and the government and around 60 meetings with senior ministers in 2018, before the final approval.[56] The powerful influence of the mining lobby led to governmental support of the project. Furthermore, Adani, mining lobbyists, and many state officials propagated false messages on the benefits of the mine, which claimed it would create 10,000 jobs despite the realistic projection of 1464 jobs according to the testimony of Adani's own economist.[57] They also disseminated false claims of "sustainable mining," despite expansive research and case studies that contradict this.[58]

In spite of the public support for decarbonization and government promises for action, the continued influence of the industry is shown in the number of mining project approvals by the Labor Government in 2024. The environmental minister approved many proposals, including 7 major mining projects, encompassing the expansion of four coal mines that will release more than 850 million tons of CO_2 over their lifetime, close to double the amount of annual national emissions.[59–62] These approvals faced public backlash and undermined Australia's goal of becoming a leader in global climate action and its bid to host COP31. Furthermore, later in the year, three more coal mine expansions were approved,[63] with the government defending its decision by saying that it is in accordance with "facts and the national environment law". However, the Australian Conservation Foundation estimated that the projects would generate 1.3 billion tons of greenhouse gas.[64] While the Liberal-National Party coalition, the conservative party, is traditionally more influenced by the mining industry, the mining expansion approvals reflect the influence it has on both sides of the political spectrum. Labor contradicted its 2022 Federal Election promises, which regarded climate change as a central issue. The outcome of the ABC Vote Compass Survey indicated that climate change is the number one priority for voters.[65] Officially, the party's climate commitments are against new mine expansions, but it is evident that policy is shifting in favor of the mining lobby.

Contrary to the previous argument on the ineffectiveness of government institutions as a factor resulting in climate policy inaction, government institutions have acted efficiently in cli-

DOI: 10.36838/IJHSR819.3

mate-related policy, especially in mining expansion approvals. Under the Environmental Protection and Biodiversity Conservation Act 1999, mining expansions that affect important national environmental outcomes, as stated in Matters of National Environmental Significance, such as water resources that relate to unconventional gas development and major coal mines, or threatened species and ecological communities, require approval from the Department of Climate Change, Energy, the Environment and Water (DCCEEW).[66] In recent years, there have been a large number of approvals by the DCCEEW, through the Minister of Water and Environment. This demonstrates the efficiency of these institutions in passing climate-related policies. However, these approvals undermine Australia's net-zero goals and are anti-climate.

In 2025, the Labor Government abandoned 'nature positive' laws that would create a government agency to enforce decisions made under environmental laws. The agency would be able to stop mining activities and audit businesses to ensure compliance with laws. Due to pressure from the mining industry before the Western Australian State Elections, these laws were not introduced because of objections from the state government. The environmental spokesperson of the Australian Greens Party, Sarah Hanson-Young, highlights that "the prime minister has again caved to the vested interests of the Western Australian mining industry."[62] These actions directly contradicted Labor election promises, with the proposed agency being a key promise of the 2022 Federal Election.[67]

Conclusion

As the world is transitioning to a green future powered by renewable energy and sustainable development, global efforts are slowly phasing out high-emitting sectors, including mining. The mining industry in Australia has employed significant amounts of resources, through financial expenditure, in order to influence major political parties through lobbies and media campaigns. The mining industry has played a big role in preventing meaningful and sustainable national policies and has hindered Australia's transition away from a heavy carbon economy. Australia has continued to be a leader on the global stage in combating anthropogenic climate change, but the influence of the mining industry undermines the government's credibility. If this continues, Australia will be unable to address its climate goals and face global backlash by breaching the legally binding Paris Agreement. Moreover, Australia's extractivist economic model, characterized by the over-reliance on finite resources in an increasingly decarbonizing world, risks its long-term development as many other nations are transitioning their economies to a sustainable model. One limitation of this research is that it does not make use of originally collected data.

In future research, it can be investigated why non-profit organizations and climate activists have been overshadowed by the interests of the mining industry. Another research study could look at the role of supranational organizations in shifting climate discourse and the national climate agenda. Moreover, the mobilization of social media on the campaign against mining expansion can be further examined. Finally, the rise of the far-right One Nation party threatens to erode the political influence of the current conservative coalition, which could change the current relationship with the mining industry. This can profoundly alter the current political landscape of Australia, leading to potential shifts in the attitudes of the government on climate policy. This might further intensify the influence of the mining industry in the future.

Acknowledgments

I would like to thank Mijoo Shin, a PhD candidate at Harvard University in Government, for her mentorship. I would also like to thank Indigo Research for their guidance.

References

1. UNFCCC. *The Paris Agreement*; Framework Convention on Climate Change. https://unfccc.int/process-and-meetings/the-paris-agreement
2. *Australian Government Climate Change commitments, policies, and programs*; Australian Office of Financial Management, 2022. https://www.aofm.gov.au/sites/default/files/2022-11-28/Aust%20Govt%20CC%20Actions%20Update%20November%202022_1.pdf
3. Climate Action Tracker. *Australia*. https://climateactiontracker.org/countries/australia/
4. Human Rights Watch. *Declaration on Mining and the Energy Transition for COP26*; Human Rights Watch, 2021. https://www.hrw.org/news/2021/10/28/declaration-mining-and-energy-transition-cop26
5. Morison, E. *Climate of the Nation 2023: Tracking Australia's attitudes towards climate change and energy*; The Australia Institute, 2023. https://australiainstitute.org.au/wp-content/uploads/2023/09/Climate-of-the-Nation-2023-Web.pdf
6. Christoff, P. Climate Discourse Complexes, National Climate Regimes and Australian Climate Policy. *Australian Journal of Politics and History* **2013**, *59* (3), 349–367, https://doi.org/10.1111/ajph.12020
7. Crowley, K. Fighting the future: The politics of climate policy failure in Australia (2015–2020). *WIREs Climate Change* **2021**, *12* (5), e725, https://doi.org/10.1002/wcc.725
8. Gascoigne, T. Climate Change: "The" Issue in the Australian National Election 2007?. *Science Communication* **2008**, *29* (4), 522–531, https://doi.org/10.1177/10755470083163
9. Cormann, M.; Turnbull, M.; Frydenburg, J. *Historical Snowy deal now complete: Joint Media Release*; Australian Government, 2018. https://ministers.finance.gov.au/financeminister/media-release/2018/07/03/historic-snowy-deal-now-complete
10. MacNeil, R. Swimming against the current: Australian climate institutions and the politics of polarisation. *Environmental Politics* **2021**, *30*, 162–183, https://doi.org/10.1080/09644016.2021.1905394
11. Snyder, J. M., Jr. Campaign Contributions as Investments: The U.S. House of Representatives, 1980-1986. *Journal of Political Economy* **1990**, *98* (6), https://doi.org/10.1086/261731
12. Australian Senate, Select Committee on the Political Influence of Donations. *Political Influence of Donations*, 2018. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Political_Influence_of_Donations/PoliticalDonations/Report_1
13. Howlett, C.; Seini, M.; McCallum, D.; Osborne, N. Neoliberalism, Mineral Development and Indigenous People: a framework for analysis. *Australian Geographer* **2011**, *42* (3), https://doi.org/10.1080/00049182.2011.595890

14. Maunder, W. J. *Australia's natural resources*; EBSCO, 2024. https://www.ebsco.com/research-starters/environmental-sciences/australias-natural-resources
15. Hyongdoo, J.; Topal, E. Transformation of the Australian mining industry and future prospects. *Mining Technology: Transactions of the Institutions of Mining and Metallurgy* **2020**, *129* (3), https://doi.org/10.1080/25726668.2020.17862
16. Connolly, E.; Orsmond, D. *The Mining Industry: From Bust to Boom*; Research Discussion Paper 2011-08; Reserve Bank of Australia, 2011. https://www.rba.gov.au/publications/rdp/2011/pdf/rdp2011-08.pdf
17. *Australia Country Commercial Guide - Mining*; United States Department of Commerce, International Trade Administration. https://www.trade.gov/country-commercial-guides/australia-mining
18. Reserve Bank of Australia. *Composition of the Australian Economy*; Reserve Bank of Australia, 2025. https://www.rba.gov.au/education/resources/snapshots/economy-composition-snapshot/
19. Australian Bureau of Statistics. *Industry profits steady in 2022-23: Media Release*; Australian Bureau of Statistics, 2024. https://www.abs.gov.au/media-centre/media-releases/industry-profits-steady-2022-23
20. Jobs and Skills Australia. *Mining*; Australian Government, 2025. https://www.jobsandskills.gov.au/data/occupation-and-industry-profiles/industries/mining
21. Dudka, S. Environmental Impacts of Metal Ore Mining and Processing: A Review. *Journal of Environmental Quality* **1997**, *26* (3), 590–602, https://doi.org/10.2134/jeq1997.00472425002600030003x
22. Punia, A.; Singh, S. K. Chapter 1 - Contamination of water resources in the mining region. *Contamination of Water: Health Risk Assessment and Treatment Strategies*; Ahamad, A.; Siddiqui, S. I.; Singh, P, Eds.; Academic Press, 2021; pp 3–17. https://doi.org/10.1016/B978-0-12-824058-8.00015-3
23. Serbula, S. M.; Milosavljevic, J. S.; Radojevic, A. A.; Kalinovic, J. V.; Kalinovic, T. S. Extreme air pollution with contaminants originating from the mining–metallurgical processes. *Science of the Total Environment* **2017**, *586*, 1066–1075, https://doi.org/10.1016/j.scitotenv.2017.02.091
24. Pandey, B.; Gautam, M.; Agrawal, M. Chapter 10 - Greenhouse Gas Emissions From Coal Mining Activities and Their Possible Mitigation Strategies. *Environmental Carbon Footprints: Industrial Case Studies*; Muthu, S. S., Ed.; Butterworth Heinemann, 2018; pp 259–294, https://doi.org/10.1016/B978-0-12-812849-7.00010-6
25. Australian Associated Press. Peter Dutton declared the Coalition government would be the mining sector's 'best friend'. *Guardian*, September 11, 2024. https://www.theguardian.com/australia-news/article/2024/sep/11/peter-dutton-declares-coalition-government-would-be-the-mining-sectors-best-friend
26. Australian Bureau of Statistics. *Australian Mining Industry, 1998-99*; Catalog No. 8414.0; Australian Bureau of Statistics, 2001. https://www.abs.gov.au/ausstats/abs@.nsf/94713ad445ff1425ca25682000192af2/93136e734ff62aa2ca2569de00271b10
27. The Australia Institute. *Mining subsidies vs public services*; The Australia Institute, 2015. https://australiainstitute.org.au/wp-content/uploads/2020/12/Brief-Mining-subsidies-vs-public-services_Final.pdf
28. The Australia Institute. *Fossil fuel subsidies*; The Australia Institute, 2025. https://australiainstitute.org.au/post/fossil-fuel-subsidies/
29. Griffiths, K.; Wood, D.; Chen, T. *How big money influenced the 2019 federal election*; Grattan Institute, 2020. https://grattan.edu.au/news/how-big-money-influenced-the-2019-federal-election/
30. The Center for Public Integrity. *Political Donations and the resource sector's influence*; The Center for Public Integrity, 2021. https://publicintegrity.org.au/political-donations-and-the-resources-sectors-influence/
31. McKay, A. What do campaign contributions buy? Lobbyists' strategic giving. *Interest Groups & Advocacy* **2018**, *7*, 1–18, https://doi.org/10.1057/s41309-018-0031-7
32. Aulby, H. *The tip of the iceberg: Political donations from the mining industry*; The Australia Institute, 2017. https://australiainstitute.org.au/wp-content/uploads/2020/12/P339-Tip-of-the-iceberg.pdf
33. Tello, E.; Hazelton, J.; Leong, S. V. Australian corporate political donation disclosures: Frequency, quality, and characteristics associated with disclosing companies. *Accounting, Auditing & Accountability Journal* **2019**, *32* (2), 581–611, https://doi.org/10.1108/AAAJ-04-2016-2515
34. Australian Electoral Commission. Disclosure Threshold; Australian Electoral Commission, 2025. https://www.aec.gov.au/parties_and_representatives/public_funding/threshold.htm
35. Griffiths, K.; Baldwin, E. *Labor is in power – but the Coalition still attracts the most money*; Grattan Institute, 2024. https://grattan.edu.au/news/labor-is-in-power-but-the-coalition-still-attracts-the-most-money/
36. Stanley, S.; Jylhä, Kirsti.; Leviston, Z.; Walker, I. Is conservative opposition to climate change threat-based? Articulating an integrated threat model of climate change attitudes. *British Journal of Social Psychology* **2023**, *63* (2), 879–893
37. *The Price of Democracy: how much do elections cost and who pays?*; Parliament of Australia, 2022. https://www.aph.gov.au/About_Parliament/Parliamentary_departments/Parliamentary_Library/Research/FlagPost/2022/May/The_price_of_democracy
38. Parliament of Australia. *Elections*. Parliamentary Handbook Online, Parliament of Australia. https://handbook.aph.gov.au/voting/elections
39. YouGov. *2025 Australian Election*; YouGov, 2025. https://au.yougov.com/elections/au/2025
40. Australian Electoral Commission. *2025 Federal Election: House of Representatives - final results*; Australian Electoral Commission, 2025. https://results.aec.gov.au/31496/Website/HouseDefault-31496.htm
41. Commonwealth Scientific and Industrial Research Organization. *Australian attitudes toward the renewable energy transition*; Commonwealth Scientific and Industrial Research Organization, 2024. https://www.csiro.au/en/research/environmental-impacts/decarbonisation/energy-transition
42. Australasian Center for Corporate Responsibility. *Advertising tricks of the fossil fuel sector*; Australasian Center for Corporate Responsibility, 2022. https://hub.accr.org.au/scoop/advertising-tricks-of-the-fossil-fuel-sector/
43. Lauder, J. Australia just approved Woodside's gas project until 2070. How could it happen? *ABC News (Australia)*, May 31, 2025. https://www.abc.net.au/news/2025-05-31/woodside-north-west-shelf-approval-did-not-consider-climate/105347716
44. Kelly, C. 'Misleading': Gina Rinehart's mining firm breached environmental code with 'clean gas' job ad, panel rules. *Guardian*, February 21, 2025. https://www.theguardian.com/media/2025/feb/21/gina-rinehart-hancock-prospecting-breached-environmental-code-with-clean-gas-job-ad-panel-rules-ntwnfb
45. *Coal Export Revenue up 186% to $112 Billion as Parliament Considers Price Caps*. The Australia Institute, 2022. https://australiainstitute.org.au/post/coal-export-revenue-up-186-to-112-billion-as-parliament-considers-price-caps/
46. Karp, P. Mining industry threatens to unleash ad campaign against Labor unless it rules out windfall profits tax. *Guardian*, Novem-

DOI: 10.36838/IJHSR819.3

ber 17, 2022. https://www.theguardian.com/australia-news/2022/nov/17/mining-industry-threatens-to-unleash-ad-campaign-against-labor-unless-it-rules-out-windfall-profits-tax
47. Staples, J. Australian Government Action in the 1980s. *Climate Change On for Young & Old*; Sykes, H., Ed.; Future Leaders, 2009; pp 112–118
48. Pearse, G. *High and Dry*; Penguin Books Australia, 2007.
49. Australian Senate, Senate Standing Committees on Environment and Communications. *Kyoto Protocol Ratification Bill 2003 [No. 2]*, 2004. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Environment_and_Communications/Completed_inquiries/2002-04/kyoto/report/index.htm
50. *Australia's Future Tax System Review Final Report*; Australian Government, The Treasury, 2010. https://treasury.gov.au/review/the-australias-future-tax-system-review/final-report
51. Hobbs, M. J. Conflict ecology: Examining the strategies and rationales of lobbyists in the mining and energy industries in Australia. *Public Relations Review* **2020**, *46* (2), https://doi.org/10.1016/j.pubrev.2019.101868
52. Bell, S.; Hindmoor, A. The Structural Power of Business and the Power of Ideas: The Strange Case of the Australian Mining Tax. *New Political Economy* **2014**, *19* (3), 470–486, https://doi.org/10.1080/13563467.2013.796452
53. Australian Bureau of Statistics. *Mining Statistics Newsletter, 2009 to 2010*; Catalog No. 8418.0; Australian Bureau of Statistics, 2010. https://www.abs.gov.au/AUSSTATS/abs@.Nsf/39433889d406eeb9ca2570610019e9a5/29843918ee9bf222ca2575d2001ae12d!OpenDocument
54. Leyton-Flor, S. A.; Sangha, K. The socio-ecological impacts of mining on the well-being of Indigenous Australians: A systematic review. *The Extractive Industries and Society* **2024**, *17*, https://doi.org/10.1016/j.exis.2024.101429
55. Lyons, K. Securing territory for mining when Traditional Owners say 'No': The exceptional case of Wangan and Jagalingou in Australia. *The Extractive Industries and Society* **2019**, *6* (3), 756–766, https://doi.org/10.1016/j.exis.2018.11.007
56. Wood, D.; Griffiths, K. *Who's in the room? Access and influence in Australian politics*; Grattan Institute, 2018. https://grattan.edu.au/wp-content/uploads/2018/09/908-Who-s-in-the-room-Access-and-influence-in-Australian-politics.pdf
57. *Expert evidence given to the case against the Adani coal mine at Carmichael*. The Australia Institute, 2015. https://australiainstitute.org.au/wp-content/uploads/2020/12/Media-Brief-Adani-Expert-Witness-Evidence-RC.pdf
58. Brevini, B.; Woronov, T. *Adani's post-truth push for the Carmichael mine;* The Conversation, 2017. https://theconversation.com/adani-is-post-truth-push-for-the-carmichael-mine-85671
59. Australian Government Department of Climate Change, Energy, the Environment and Water. *Caval Ridge Mine Horse Pit Extension, Bowen Basin*; Environmental Protection and Biodiversity Conservation Act Public Portal, 2024. https://epbcpublicportal.environment.gov.au/all-notices/project-decision/?id=6b1c7381-2c53-ed11-9562-00224818a6e7
60. Australian Government Department of Climate Change, Energy, the Environment and Water. *Lake Vermont Meadowbrook Coal Mine Project, Qld*; Environmental Protection and Biodiversity Conservation Act Public Portal, 2024. https://epbcpublicportal.environment.gov.au/all-notices/project-decision/?id=43ea38df-6df8-ee11-a1fe-00224897164a
61. Australian Government Department of Climate Change, Energy, the Environment and Water. *Boggabri Coal Mine SSD 09_0182 Modification 8*; Environmental Protection and Biodiversity Conservation Act Public Portal, 2024. https://epbcpublicportal.environment.gov.au/all-notices/project-decision/?id=9c23be9c-f1d5-ee11-904c-6045bde708a5
62. Australian Government Department of Climate Change, Energy, the Environment and Water. *Vulcan South Coal Mine*; Environmental Protection and Biodiversity Conservation Act Public Portal, 2024. https://epbcpublicportal.environment.gov.au/all-referrals/project-referral-summary/?id=4ae90f96-9582-ee11-8179-000d3a794b99
63. Roe, I.; Lowrey, T. Government green-lights three NSW coal mine extensions, angering environmental groups. *ABC News (Australia)*, September 24, 2024. https://www.abc.net.au/news/2024-09-24/federal-government-approves-coal-mine-extensions/104391416
64. Meadows, J. *Coal approvals: not a good look and disastrous for climate and nature*; Australian Conservation Foundation, 2024. https://www.acf.org.au/news/coal-approvals-not-a-good-look-and-disastrous-for-climate-and-nature
65. Slezak, M. Federal election 2022: Labor and Coalition climate change policies both breach the Paris Agreement. So what does the science say?. *ABC News (Australia)*, May 18, 2022. https://www.abc.net.au/news/2022-05-18/federal-election-climate-change-policy-labor-liberal-greens/101074924
66. *Environmental Protection and Biodiversity Conservation Act 1999* (Australia).
67. Speers, D.; Truu, M. Labor shelves contentious 'nature positive' laws after West Australian backlash. *ABC News (Australia)*, February 2, 2025. https://www.abc.net.au/news/2025-02-02/labor-shelves-contentious-nature-positive-laws/104886608

Author

Andy Guan is a junior at Redlands School, Australia. He is passionate about international relations, global affairs, and climate politics, particularly on how history has shaped the modern world. In the future, he wants to study International Relations at university.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

LGI1 in Epilepsy: Genetic Synaptopathies and Autoimmune Mechanisms

Cara M. de Wesselow

Tanglin Trust School, 95 Portsdown Road, 139299, Singapore; cmdewesselow@icloud.com

ABSTRACT: Epilepsy is a devastating neurological condition affecting an estimated 50 million individuals worldwide, arising from neuronal hyperexcitability due to disrupted excitatory-inhibitory balance. Mutations in the LGI1 (leucine-rich glioma inactivated 1) gene are the most common cause of Autosomal Dominant Lateral Temporal Lobe Epilepsy (ADLTE), a condition characterized by auditory and aphasic auras, often progressing into focal seizures. Moreover, the production of anti-LGI1 antibodies, inhibiting LGI1-ADAM22/23 interaction through blocked binding sites, causes anti-LGI1 Autoimmune Encephalitis, a distinct AE subtype. Patients with this condition suffer from memory disturbance, seizures, and may exhibit psychiatric symptoms (e.g., hallucinations, delirium). These pathophysiological mechanisms of LGI1 dysfunction (genetic and autoimmune) both give rise to epileptic conditions. Thus, it is a significant protein in understanding the molecular biology behind epilepsy and may have implications for the future of diagnosis and treatment of epilepsy. This literature review focuses on the structure, physiological function, and seizure-related disruption of LGI1, which regulates AMPA-R clustering through LGI1-ADAM22/23 and PSD95 interaction, linking presynaptic and postsynaptic proteins to stabilize excitatory synaptic transmission.

KEYWORDS: Medical and Health Sciences, Neurobiology, Epilepsy, Synaptic Transmission, LGI1.

Introduction

Epilepsy is one of the oldest and most common neurological conditions known to mankind, affecting individuals of both sexes and all ages.[1,2] The World Health Organization estimates that 50 million individuals worldwide have a diagnosis of epilepsy at any given time.[3] It is a chronic disorder characterized by recurrent, unprovoked seizures; seizures are alterations of neurologic function caused by the excessive, hypersynchronous discharge of neurons in the brain.[4,5] Symptoms of epilepsy include loss of consciousness, anxiety, confusion, staring spells, muscle jerking/twitching, and muscle stiffness, as shown in Figure 1.[6]

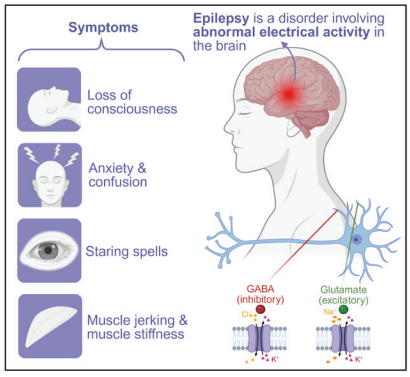


Figure 1: Epilepsy symptoms and neuronal role. Schematic illustration depicting a focal region of abnormal neuronal activity within the brain, with a simplified representation of a neuron highlighting disrupted excitatory-inhibitory signaling. Excessive glutamatergic excitation and reduced GABAergic inhibition increase neuronal firing through altered ion fluxes (e.g., K^+, Cl^-, Na^+), leading to hyperexcitability and seizures. Symptoms include: loss of consciousness and confusion arising from disruption across many brain regions, staring spells indicating impairments of awareness, and involuntary muscle movement arising from abnormal neuronal activity. Created using biorender.com.

As depicted in Figure 1, normal neuronal function involves receiving a mix of transmitted signals. Glutamate (released at excitatory synapses) binds to receptors and increases the likelihood of the neuron firing an action potential, whereas GABA (released at inhibitory synapses) binds to receptors and decreases the likelihood of the neuron firing an action potential.[7,8] The excitatory/inhibitory balance determines whether a neuron becomes active or remains silenced; in epilepsy, seizures are caused by hyperexcitability stemming from this imbalance, causing groups of neurons to become hyperexcitable and fire in a hypersynchronous manner, producing sustained abnormal discharges in the brain.[9,10]

Most epilepsy pathophysiology mechanisms can be grouped into channelopathies, synaptopathies, structural abnormalities of the brain, and autoimmune mechanisms. Channelopathies, or ion channel dysfunction, are caused by mutations in voltage-gated or ligand-gated ion channels, which alter neuronal excitability, leading to repetitive firing.[11,12] Synaptopathies, or synaptic protein dysfunction, involve abnormalities in proteins that govern synapse regulation, altering neuronal excitability and disrupting the excitatory/inhibitory balance.[13] Structural abnormalities causing epilepsy include malformations of cortical development,[14] or acquired causes such as brain tumors or trauma. Autoimmune mechanisms arise due to autoantibodies against synaptic proteins (e.g. anti-α-amino-3-hydroxy-5-methyl-4- isoxazolepropionic-acid receptor (anti-AMPAR), anti-N-methyl-D-aspartate receptor (NMDAR)), which cause protein dysfunction, leading to hyperexcitability in autoimmune encephalitis.[15,16]

Another way of categorizing epilepsy cases is by seizure type. There are three major categories of epileptic seizure types classified by the International League Against Epilepsy (ILAE),

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.4

most recently revised in 2017.[17] The first category is epilepsy characterized by focal (partial) seizures, which have a specific origin in the brain; they can be further subdivided into focal aware seizures, in which consciousness is preserved, and focal impaired awareness seizures, in which consciousness is disrupted or lost.[18,19] The second category is epilepsy, characterized by generalized seizures, which occur in both hemispheres of the brain simultaneously.[20] The third category involves seizures of unknown onset, if there is insufficient information to classify them.[17]

To understand the molecular biology of epilepsy, it is important to analyze different mechanistic categories of epilepsy pathophysiology, especially through the lens of proteins whose dysfunction gives rise to epileptic phenotypes. One such protein is LGI1, which is particularly relevant to epilepsy research due to its dual vulnerability to two mechanistic categories of epilepsy pathophysiology mentioned above: synaptopathies and autoimmune disorders. In Autosomal Dominant Lateral Temporal lobe Epilepsy (ADLTE), inherited mutations in the LGI1 gene impair excitatory-inhibitory balance and synaptic maturation.[21] In anti-LGI1 Autoimmune Encephalitis (AE), antibodies disrupt LGI1's extracellular interactions.[22] Despite their distinct origins (genetic versus autoimmune), the mechanisms causing ADLTE and anti-LGI1 AE affect the same proteins, and both produce epileptic phenotypes. Therefore, LGI1 is shown to have a central role in maintaining synaptic integrity and to be suitable as a model of synaptic vulnerability. This review analyzes the molecular mechanisms underlying ADLTE and anti-LGI1 AE through the LGI1 network, and discusses the implications of this analysis on our understanding, diagnosis, and treatment of epilepsy.

Discussion

LGI1 (leucine-rich, glioma inactivated 1) has emerged as a particularly suitable model of synaptic vulnerability in epilepsy. LGI1 is uniquely positioned at the excitatory synapse, where it forms trans-synaptic complexes to maintain the stability of synaptic transmission. Disruption of this role is linked to epileptic phenotypes exhibited in ADLTE and anti-LGI1 Autoimmune Encephalitis. Therefore, a thorough analysis of LGI1 provides an opportunity to examine the relationship between synaptic dysfunction and epilepsy.

LGI1-mediated protein interaction at the synapse:

LGI1 (leucine-rich, glioma inactivated 1) is a secreted neuronal synaptic protein that plays key roles in excitatory neurotransmission, synaptic maturation, and maintaining network stability.[23] In a study, purification of neuronal LGI1 led to the isolation of abundant components of the LGI1-mediated protein complex. The relevant proteins in understanding LGI1 protein network function include ADAM22 and other ADAM family members (ADAM23, 11) and Post Synaptic Density Proteins (PSD95, PSD93).[24]

ADAM22 is associated with the PSD95 complex found in excitatory postsynaptic densities; ADAM22 has a short C-terminal intracellular sequence that anchors it to PSD95 through its PDZ domain.[25] PSD95 anchors AMPA receptors by interacting with Stargazin, a TARP (transmembrane AMPA-R regulatory protein).[22,26] The LGI1-ADAM22 complex stabilizes AMPA receptors by reinforcing their connection to PSD95, making it essential for efficient regulation of excitatory neurotransmission. It has been shown that the LGI1-ADAM22 complex forms a tether that aligns presynaptic glutamate release sites with postsynaptic AMPA receptors, proving it is fundamental for transsynaptic nanoalignment, as shown in Figure 2.[27] Although LGI1 has been proven to bind ADAM23 through ELISA tests and transfected HEK cells, there is still uncertainty surrounding their direct interaction and physiological function.[28,29] However, ADAM23 is a major receptor of LGI1 and is likely to play an important role, presumably in regulating neurotransmission.[30]

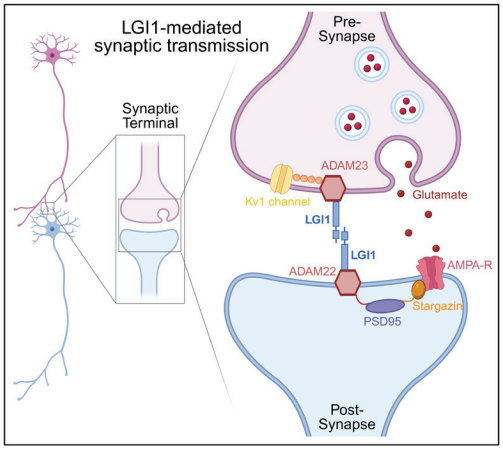


Figure 2: LGI1 in the trans-synaptic space. Schematic depicting transmission between excitatory neurons (from nerve endings of one neuron to dendrites of another) at the synaptic terminal. Calcium ions flood into the presynaptic terminal, causing vesicles to fuse with the membrane and release glutamate neurotransmitters into the synaptic cleft. Glutamate diffuses across the cleft and binds to AMPA receptors, which in turn causes an action potential in the postsynaptic neuron. LGI1 mediates this process by binding ADAM22 (postsynaptic) and ADAM23 (presynaptic), forming a transsynaptic tether. As ADAM22 interacts with the PSD95 scaffolding protein, the LGI1-ADAM22 complex stabilizes and clusters AMPA receptors by reinforcing their connection to PSD95 (which indirectly interacts with AMPA receptors through stargazin, a TARP). For the sake of simplicity, this diagram does not depict the complex 3:3 heterohexameric LGI1-ADAM22 assembly, showing only the interaction of LGI1 molecules with ADAM22/23. Created using biorender.com.

LGI1 protein structure:

LGI1 is a glycoprotein with a molecular mass of approximately 60-64 kDa.[27,29] LGI1 is a member of the leucine-rich glioma inactivated family, which consists of secreted glycoproteins with a leucine-rich repeat domain, or LRR, and an epitempin domain, or EPTP, also known as the epilepsy-associated repeat domain (EAR).[31] A computational rendering of the structure described below is depicted in Figure 3. The LRR domain contains four LRR repeats, flanked by LRRNT (leucine-rich repeat N-terminal, the start of the chain) and LRRCT (leucine-rich repeat C-terminal, the end of the chain).[32] Both the N-terminal and C-terminal are cysteine-rich and form intramolecular disulfide bonds that 'cap' the ends of the LRR solenoid, reinforcing its structural integrity, which is crucial to LGI1's role as a secreted protein.[33] The EPTP (epitempin) domain has seven EPTP repeats,

which fold into a structure that resembles a conical doughnut, known as a seven-bladed β-propeller; each blade is composed of a four-stranded antiparallel β-sheet.[34] Notable features of LGI1 EPTP are the disulfide bond inside the β-propeller and a Ca^{2+} ion bound inside the central channel, which stabilize the whole β-propeller structure and ensure correct orientation.[35] The name epitempin comes from the observation that this domain is found in LGI1 and another epilepsy-associated protein, GPR98 (a G-protein-coupled receptor).[36] Mutations in epitempin-encoding genes LGI1 and GPR98 both produce auditory symptoms such as noise-induced seizures or deafness.[37]

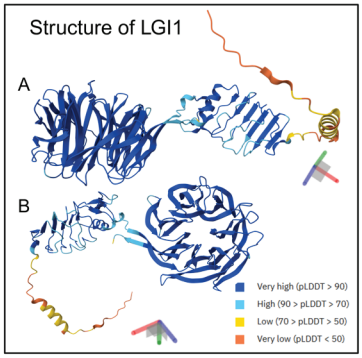


Figure 3: LGI1 protein structure. (A) Computational rendering of the LGI1 structure shows a sideways view of the EPTP domain β-propeller structure; β-pleated sheets are visible. LGI1 predominantly features β-pleated sheets in both its EPTP and LRR domains; it features minimal to no α-helices. (B) Another orientation of the LGI1 protein is shown, including a front view of the β-propeller structure consisting of seven four-stranded antiparallel β-sheets folded into a doughnut-like shape, as well as the LRR domain and its β-pleated sheet composition. Source: alphafold.com.

The LGI1-ADAM22 complex:

ADAM family proteins consist of multiple domains, including a metalloproteinase domain, a disintegrin domain, a transmembrane domain, a cytoplasmic carboxy (C)-terminal tail, and an epidermal growth factor (EGF)-like domain.[29,38] Eight of twenty-one ADAMs lack functional zinc metalloproteinase domains and are instead associated with protein-protein interactions instead of membrane protein ectodomain shedding (a type of post-translational modification).[39] Three ADAM proteins (ADAM22, ADAM23, ADAM11) bind to LGI1 and are solely expressed in the central nervous system.[40] ADAM22, the first major receptor of LGI1, has an intracellular domain that interacts with PSD-95 and AMPA receptors in the postsynaptic density, and an extracellular domain, or ECD, which interacts with LGI1 to govern transsynaptic nanoalignment.[41,42] The seven-bladed β-propeller of LGI1 forms a hydrophobic pocket, and the metalloprotease-like domain of ADAM22 inserts into the hydrophobic pocket, forming a tight link, but without covalent bonding, as shown in Figure 4A.[35,44]

Structural analysis through crystallization of the LGI1-ADAM22 complex revealed a 2:2 heterotetrameric assembly, meaning two LGI1 molecules and two ADAM22 molecules form a four-part complex, where each LGI1 binds to one ADAM22, and LGI1 molecules interact with each other.[27] In this way, it was proposed that LGI1 bridges two ADAM22 molecules through LRR-EPTP interaction of adjacent LGI1 molecules.[35] This 2:2 tetrameric configuration was further supported using cryo-electron microscopy analysis.[43] However, recent cryo-electron microscopy analyses have suggested the existence of a 3:3 heterohexameric higher-order complex, which forms a ring-like structure (where each LGI1 bridges two ADAM22 molecules), depicted in Figure 4B.[27,44,45] This evolution in structural understanding suggests a more stable and functionally relevant synaptic architecture.

ADAM23, the second major receptor of LGI1, has an intracellular domain in the presynaptic density and an extracellular domain that interacts with LGI1. It likely interacts with LGI1 in a manner similar to ADAM22, since the amino acid sequence similarity between ADAM22 and ADAM23 is relatively high in humans (approx. 50%).[27,46]

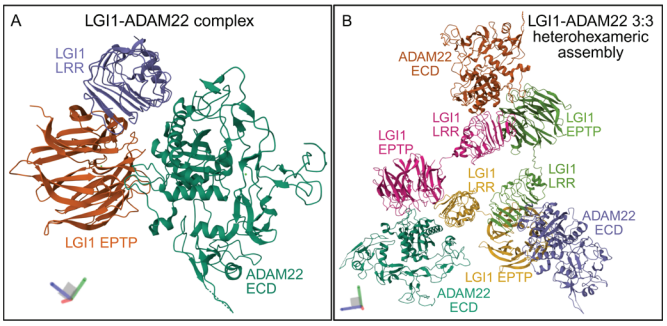


Figure 4: LGI1-ADAM22 complex and 3:3 heterohexameric assembly. (A) The metalloprotease-like domain (ECD) of ADAM22 interacts directly with LGI1 EPTP by inserting into the hydrophobic pocket created by the seven-bladed β-propeller. LGI1 LRR interacts weakly/indirectly with ADAM22 ECD, providing structural support. (B) Three LGI1 molecules associated with three ADAM22 molecules in a symmetrical ring-like arrangement, forming a 3:3 heterohexameric higher-order assembly. The assembly stabilizes the LGI1-ADAM22 complex and allows it to perform its functions (organizing AMPA receptors via ADAM22-PSD95 interactions, etc.). Structure determined experimentally by cryo-electron microscopy (cryo-EM). Source: RCSB Protein Data Bank.

LGI1 pathogenic mutations in ADLTE:

Autosomal dominant lateral temporal lobe epilepsy (ADLTE), also known as autosomal dominant partial epilepsy with auditory features (ADPEAF), is a genetic focal epilepsy syndrome characterized by recurrent seizures with aura suggestive of a lateral temporal lobe onset.[35,47,48] The lateral temporal lobe (especially the superior temporal gyrus and middle temporal gyrus) houses primary and association auditory cortex and the language comprehension area. Therefore, ADLTE symptoms include auditory auras and aphasic auras, often progressing to focal seizures. Visual and olfactory symptoms have also been reported.[49,50]

Interestingly, déjà vu, defined by V.M. Neppe in 1983 as "any subjectively inappropriate impression of familiarity of a present experience with an undefined past", has been reported as a predominant feature of seizures in ADLTE patients in several cases.[51,52] The explanation can be inferred by an experiment done by Halgren in 1978, who proved that a 'dreamy state', consisting of vivid memory-like hallucinations and déjà vu, can be evoked by electrical stimulation of the lateral temporal neocortex, as a result of spontaneous seizures.[53]

DOI: 10.36838/IJHSR819.4

ADLTE is associated with inherited mutations in the LGI1, RELN, and MICAL1 genes.[54] This literature review focuses on the primary, most common cause of ADLTE, which is LGI1 and its pathogenic mutations. Several secretion-positive mutations, including R474Q and R407C, impair LGI1's ability to bind to neuronal receptors ADAM22/23 by altering its EPTP domain, which is crucial for receptor binding.[55] The R474Q mutation disables the tripartite complex assembly of ADAM22, ADAM23 and LGI1 by disrupting LGI1-LGI1 interaction. Many secretion-defective mutations cause protein misfolding, including A110D, I122K, and R136W. Other secretion-defective mutations cause disruption of disulfide bonds, including C46R (disruption of the disulfide bond with C55) and C200R (disruption of the disulfide bond with C177).[24,27,35] These mutations, among others, are named and organized by location in Figure 5.

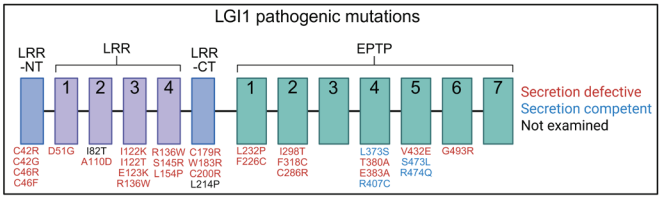


Figure 5: Pathogenic LGI1 mutations. Diagram showing known pathogenic LGI1 mutations, separated into secretion-defective (causing misfolding or disruption of disulfide bonds), secretion-competent (LGI1 is secreted, but there is disruption of binding to ADAM22 or other LGI1 molecules), and not experimentally examined. The mutations are organized by the location on LGI1 they affect; the LRR domain is split into the N-terminal, C-terminal, and 4 LRR repeats, & the EPTP domain is split into the 7 EPTP repeats. As shown, all secretion-competent mutations known affect the EPTP domain. Created using biorender.com.

Pathogenic LGI1 mutations, regardless of secretion status, disrupt LGI1's physiological function, which is to bind ADAM22/23, forming a transsynaptic tether.[29] Secretion-defective mutations cause LGI1 to misfold and be trapped in the endoplasmic reticulum, unable to be secreted into the synaptic cleft to bind ADAM22/23. Secretion-competent mutations prevent LGI1 from properly binding to ADAM22/23, breaking the transsynaptic complex.[56] This causes AMPA-R clustering to be disturbed, and the excitatory/inhibitory balance is not fully maintained, leaning towards hyperexcitability (especially in the lateral temporal cortex, a region where auditory and association processing is abundant).[57,58]

Beyond its central role in maintaining AMPA receptor clustering in mature synapses, LGI1 is also a regulator of synaptic maturation and participates in broader developmental processes.[59] In murine models, LGI1 mRNA and protein are expressed during embryonic brain development, and expression persists and often increases postnatally. Studies emphasize its functional importance during the postnatal period, suggesting its role in synaptic maturation.[60,61] Expression of LGI1 during development follows an orchestrated pattern and is regionally specific, particularly in cortical and temporal regions (which show seizure susceptibility).[62] This implies that LGI1 may be involved in shaping brain architecture and neuronal function during brain development. Therefore, in ADLTE, pathogenic LGI1 mutations are likely to disrupt not only synaptic stabilization at mature excitatory synapses but also synaptic maturation, leading to hyperexcitability in regions susceptible to epilepsy, such as the lateral temporal cortex.

Anti-LGI1 autoimmune encephalitis:

Autoimmune encephalitis (AE) comprises a group of inflammatory brain disorders characterized by acute or subacute onset of symptoms, including seizures, memory impairment, altered consciousness, and behavioral change.[63,64] They are caused by immune responses against neuronal antigens, often autoantibodies against antigens localized on the neuronal cell surface or at synapses, leading primarily to functional neuronal disruption (as opposed to structural disruption).[65]

Anti-LGI1 antibody-associated autoimmune encephalitis, an AE subtype, is the second most prevalent autoimmune encephalitis after anti-N-methyl-D-aspartate receptor (NMDAR) encephalitis, accounting for 20-35% of seropositive cases.[66] Other AE subtypes include anti-NMDAR, anti-CASPR2, anti-GABAR, and anti-AMPAR.[67] Patients with anti-LGI1 antibodies develop symptoms of limbic encephalitis, manifesting in memory disturbance, distinct seizure types, confusion, etc. Patients may also show psychiatric symptoms such as hallucinations, delusions, and delirium with disorganized speech.[67]

At the molecular level, LGI1 antibodies associated with limbic encephalitis inhibit LGI1-ADAM22/23 interaction by targeting the EPTP domain of LGI1, greatly reducing synaptic AMPA receptor clusters.[68] Reduced LGI1-ADAM22/23 interaction is the most reasonable pathogenic mechanism for limbic encephalitis. This is supported by the fact that genetic deletion of LGI1 or ADAM22/23 causes lethal epilepsy in murine models and is further backed up by the presence of secretion-defective LGI1 mutations in ADLTE.[30] Although knowledge on the precise mode of action of LGI1 antibodies remains incomplete, synaptic autoimmune disorders are thought to involve antibody-mediated dysfunction of target ion channels and key synaptic proteins, such as NMDAR and AMPAR, leading to network hyperexcitability, causing seizures.[69]

■ Conclusion

Epilepsy is a common neurological condition caused by the disruption of excitatory-inhibitory neuronal balance (hyperexcitability), leading to seizures. As previously discussed, LGI1 is a fundamental regulator of synaptic transmission, as LGI1-mediated interactions with ADAM22/23 and PSD95 are essential in AMPA-R clustering; it is uniquely positioned to act as a transsynaptic 'tether', linking presynaptic and postsynaptic proteins. ADLTE and anti-LGI AE represent two distinct mechanisms causing epilepsy (one genetic and one autoimmune) that converge on the same molecular pathway involving dysfunction of LGI1-mediated synaptic stabilization. In ADLTE, pathogenic mutations impair the secretion of LGI1 or its ability to form a functional complex with ADAM22/23, whereas in anti-LGI1 AE, autoantibodies block this interaction at the synapse. Another significant difference is that ADLTE is a developmental disorder arising from abnormalities during brain development, while anti-LGI1 AE is an

acquired disorder that develops after normal brain development (as a result of autoantibodies). However, both conditions destabilize AMPAR clustering and disrupt excitatory neurotransmission, leading to hyperexcitability in the network. This similarity highlights LGI1 as a critical player in seizure susceptibility.

One diagnostic implication involves genetic testing for LGI1 mutations in ADLTE; however, in practice, genetic testing for LGI1 may not be widely performed. This may be due to the presence of other possible gene mutations causing ADLTE (such as RELN), although LGI1 mutations make up most cases. It may also be due to a diagnosis based on EEG findings and symptoms alone. Nevertheless, earlier identification of LGI1-related ADLTE may prevent misdiagnosis from finding no structural abnormalities; clinicians must understand the genetic component of ADLTE. A second diagnostic implication involves early testing for anti-LGI1 antibodies in patients with seizures, limbic symptoms, or other symptoms relevant to autoimmune encephalitis. Consideration of anti-LGI1 antibodies could prevent delayed diagnosis. Even so, anti-LGI1 antibody testing sensitivity and specificity may not be optimal.

Understanding LGI1 and its pathophysiology is highly relevant to future implications on the treatment of ADLTE and anti-LGI1 AE. Current treatments mostly suppress seizures and treat symptoms rather than correcting the underlying molecular dysfunction. For example, current treatment of ADLTE remains symptomatic, but gene editing may be a plausible treatment in the future. Another future direction could be administering drugs to stabilize synaptic transmission or moderate network excitability in ADLTE, compensating for LGI1 dysfunction. In contrast, anti-LGI1 autoimmune encephalitis is already treated, often effectively, through immunotherapy. However, it may be useful to understand anti-LGI1 autoantibody mechanisms to optimize immunotherapy in the future. Of note, because this literature review primarily focuses on LGI1 as a model of synaptic transmission in epilepsy, it focuses on excitatory synaptic dysfunction, particularly AMPA-R clustering, whereas epilepsy arises from an excitatory-inhibitory imbalance in a wider neuronal network. Alterations in inhibitory GABAergic transmission were not investigated in depth. In addition, zooming in on a single molecular pathway risks oversimplifying the complexity of the pathophysiology of epilepsy; further exploration into other mechanistic categories is needed.

In conclusion, LGI1 and its associated epilepsies illustrate how disruption of a single synaptic protein can give rise to diverse epileptic phenotypes, underscoring the significance of molecular mechanisms in understanding, diagnosis, and treatment of epilepsy. As our understanding of synaptic protein networks deepens, LGI1 may serve as a reference point for analyzing other epilepsy-associated synaptic proteins, paving the way for molecular mechanism-based diagnosis and therapies.

Acknowledgments

I would like to thank my mentor, Dr Kif Liakath-Ali, for his guidance, feedback, and support throughout the development of this literature review. This research opportunity has allowed me to gain a deeper understanding of topics within neurobiology and molecular biology, fields I am deeply passionate about.

References

1. World Health Organization. *Epilepsy*. https://www.who.int/health-topics/epilepsy (accessed 2026-02-02).
2. Banerjee, P. N.; Filippi, D.; Allen Hauser, W. The Descriptive Epidemiology of Epilepsy—A Review. *Epilepsy Res.* **2009**, *85* (1), 31–45. https://doi.org/10.1016/j.eplepsyres.2009.03.003.
3. Neligan, A.; Sander, J. W. The Incidence and Prevalence of Epilepsy.
4. Beghi, E. The Epidemiology of Epilepsy. *Neuroepidemiology* **2019**, *54* (2), 185–191. https://doi.org/10.1159/000503831.
5. Stafstrom, C. E.; Carmant, L. Seizures and Epilepsy: An Overview for Neuroscientists. *Cold Spring Harb. Perspect. Med.* **2015**, *5* (6), a022426. https://doi.org/10.1101/cshperspect.a022426.
6. Gavvala, J. R.; Schuele, S. U. Epilepsy. *JAMA* **2016**, *316* (24), 2686. https://doi.org/10.1001/jama.2016.18623.
7. Chen, T.-S.; Huang, T.-H.; Lai, M.-C.; Huang, C.-W. The Role of Glutamate Receptors in Epilepsy. *Biomedicines* **2023**, *11* (3). https://doi.org/10.3390/biomedicines11030783.
8. Sotero, R. C.; Trujillo-Barreto, N. J. Modeling the Role of Excitatory and Inhibitory Neuronal Activity in the Generation of the BOLD Signal. *Neuroimage* **2007**, *35* (1), 149–165. https://doi.org/10.1016/j.neuroimage.2006.10.027.
9. Bonansco, C.; Fuenzalida, M. Plasticity of Hippocampal Excitatory-Inhibitory Balance: Missing the Synaptic Control in the Epileptic Brain. *Neural Plast.* **2016**, *2016*, 8607038. https://doi.org/10.1155/2016/8607038.
10. Engelborghs, S.; D'hooge, R.; De Deyn, P. P. Pathophysiology of Epilepsy. *Acta Neurol. Belg.* **2000**, *100* (4), 201–213.
11. Bernard, G.; Shevell, M. I. Channelopathies: A Review. *Pediatr. Neurol.* **2008**, *38* (2), 73–85. https://doi.org/10.1016/j.pediatrneurol.2007.09.007.
12. Lerche, H.; Shah, M.; Beck, H.; Noebels, J.; Johnston, D.; Vincent, A. Ion Channels in Genetic and Acquired Forms of Epilepsy. *J. Physiol.* **2013**, *591* (4), 753–764. https://doi.org/10.1113/jphysiol.2012.240606.
13. Scharfman, H. E. The Neurobiology of Epilepsy. *Curr. Neurol. Neurosci. Rep.* **2007**, *7* (4), 348–354. https://doi.org/10.1007/s11910-007-0053-z.
14. Guerrini, R.; Dobyns, W. B. Malformations of Cortical Development: Clinical Features and Genetic Causes. *Lancet Neurol.* **2014**, *13* (7), 710–726. https://doi.org/10.1016/S1474-4422(14)70040-7.
15. Dalmau, J.; Graus, F. Antibody-Mediated Encephalitis. *N. Engl. J. Med.* **2018**, *378* (9), 840–851. https://doi.org/10.1056/NEJMra1708712.
16. Husari, K. S.; Dubey, D. Autoimmune Epilepsy. *Neurotherapeutics* **2019**, *16* (3), 685–702. https://doi.org/10.1007/s13311-019-00750-3.
17. Fisher, R. S.; Cross, J. H.; French, J. A.; Higurashi, N.; Hirsch, E.; Jansen, F. E.; Lagae, L.; Moshé, S. L.; Peltola, J.; Roulet Perez, E.; Scheffer, I. E.; Zuberi, S. M. Operational Classification of Seizure Types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. *Epilepsia* **2017**, *58* (4), 522–530. https://doi.org/10.1111/epi.13670.
18. Cook, M.; Guerrini, R.; Shorvon, S. D.; Lhatoo, S. D., Eds. *Oxford Textbook of Epilepsy and Epileptic Seizures*; Oxford University Press: Oxford, 2013.
19. Contreras Ramirez, V.; Vaddiparti, A.; Blumenfeld, H. Testing Awareness in Focal Seizures: Clinical Practice and Interpretation

DOI: 10.36838/IJHSR819.4

of Current Guidelines. *Ann. Clin. Transl. Neurol.* **2022**, *9* (5), 762–765. https://doi.org/10.1002/acn3.51552.
20. Maguire, M.; Marson, A. G.; Ramaratnam, S. Epilepsy (Generalized). *BMJ Clin. Evid.* **2012**, *2012*, 1201.
21. Lee, M. K.; Kim, S. W.; Lee, J. H.; Cho, Y.-J.; Kim, D.-E.; Lee, B. I.; Kim, H. M.; Lee, M. G.; Heo, K. A Newly Discovered LGI1 Mutation in a Korean Family with Autosomal Dominant Lateral Temporal Lobe Epilepsy. *Seizure* **2014**, *23* (1), 69–73. https://doi.org/10.1016/j.seizure.2013.10.001.
22. Fels, E.; Muñiz-Castrillo, S.; Vogrig, A.; Joubert, B.; Honnorat, J.; Pascual, O. Role of LGI1 Protein in Synaptic Transmission: From Physiology to Pathology. *Neurobiol. Dis.* **2021**, *160*, 105537. https://doi.org/10.1016/j.nbd.2021.105537.
23. Lovero, K. L.; Fukata, Y.; Granger, A. J.; Fukata, M.; Nicoll, R. A. The LGI1–ADAM22 Protein Complex Directs Synapse Maturation through Regulation of PSD-95 Function. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (30), E4129–E4137. https://doi.org/10.1073/pnas.1511910112.
24. Fukata, Y.; Fukata, M. Epilepsy and Synaptic Proteins. *Curr. Opin. Neurobiol.* **2017**, *45*, 1–8. https://doi.org/10.1016/j.conb.2017.02.001.
25. Hsia, H.-E.; Tüshaus, J.; Brummer, T.; Zheng, Y.; Scilabra, S. D.; Lichtenthaler, S. F. Functions of "A Disintegrin and Metalloproteases (ADAMs)" in the Mammalian Nervous System. *Cell. Mol. Life Sci.* **2019**, *76* (16), 3055–3081. https://doi.org/10.1007/s00018-019-03173-7.
26. Kegel, L.; Aunin, E.; Meijer, D.; Bermingham, J. R. LGI Proteins in the Nervous System. *ASN Neuro* **2013**, *5* (3), AN20120095. https://doi.org/10.1042/AN20120095.
27. Yamagata, A.; Miyazaki, Y.; Yokoi, N.; Shigematsu, H.; Sato, Y.; Goto-Ito, S.; Maeda, A.; Goto, T.; Sanbo, M.; Hirabayashi, M.; Shirouzu, M.; Fukata, Y.; Fukata, M.; Fukai, S. Structural Basis of Epilepsy-Related Ligand–Receptor Complex LGI1–ADAM22. *Nat. Commun.* **2018**, *9*, 1546. https://doi.org/10.1038/s41467-018-03947-w.
28. Sagane, K.; Ishihama, Y.; Sugimoto, H. LGI1 and LGI4 Bind to ADAM22, ADAM23 and ADAM11. *Int. J. Biol. Sci.* **2008**, *4* (6), 387–396. https://doi.org/10.7150/ijbs.4.387.
29. Baudin, P.; Cousyn, L.; Navarro, V. The LGI1 Protein: Molecular Structure, Physiological Functions and Disruption-Related Seizures. *Cell. Mol. Life Sci.* **2021**, *79*, 16. https://doi.org/10.1007/s00018-021-04088-y.
30. Owuor, K.; Harel, N. Y.; Englot, D. J.; Hisama, F.; Blumenfeld, H.; Strittmatter, S. M. LGI1-Associated Epilepsy through Altered ADAM23-Dependent Neuronal Morphology. *Mol. Cell. Neurosci.* **2009**, *42* (4), 448–457. https://doi.org/10.1016/j.mcn.2009.09.008.
31. Pakozdy, A.; Patzl, M.; Zimmermann, L.; Jokinen, T. S.; Glantschnigg, U.; Kelemen, A.; Hasegawa, D. LGI Proteins and Epilepsy in Humans and Animals. *J. Vet. Intern. Med.* **2015**, *29* (4), 997–1005. https://doi.org/10.1111/jvim.12610.
32. Somerville, R. P. T.; Chernova, O.; Liu, S.; Shoshan, Y.; Cowell, J. K. Identification of the Promoter, Genomic Structure, and Mouse Ortholog of LGI1. *Mamm. Genome* **2000**, *11* (8), 622–627. https://doi.org/10.1007/s0033500101280.
33. Ponferrada, V.; Rousseau, J.; Migliorini, G.; *et al.* Structural Analysis of Leucine-Rich-Repeat Variants in Proteins Associated with Human Diseases. *Cell. Mol. Life Sci.* [Online]. https://link.springer.com/article/10.1007/s00018-005-5187-z (accessed 2026-02-03).
34. Limviphuvadh, V.; Chua, L. L.; Rahim, R. A. B.; Eisenhaber, F.; Maurer-Stroh, S.; Adhikari, S. Similarity of Molecular Phenotype between Known Epilepsy Gene LGI1 and Disease Candidate Gene LGI2. *BMC Biochem.* **2010**, *11*, 39. https://doi.org/10.1186/1471-2091-11-39.
35. Yamagata, A.; Fukai, S. Insights into the Mechanisms of Epilepsy from Structural Biology of LGI1–ADAM22. *Cell. Mol. Life Sci.* **2020**, *77* (2), 267–274. https://doi.org/10.1007/s00018-019-03269-0.
36. Kegel, L.; Aunin, E.; Meijer, D.; Bermingham, J. R. LGI Proteins in the Nervous System. *ASN Neuro* **2013**, *5* (3), AN20120095. https://doi.org/10.1042/AN20120095.
37. Bousfiha, A.; Bakhchane, A.; Charoute, H.; Detsouli, M.; Rouba, H.; Charif, M.; Lenaers, G.; Barakat, A. Novel Compound Heterozygous Mutations in the GPR98 (USH2C) Gene Identified by Whole Exome Sequencing in a Moroccan Deaf Family. *Mol. Biol. Rep.* **2017**, *44* (5), 429–434. https://doi.org/10.1007/s11033-017-4129-9.
38. Benarroch, E. E. ADAM Proteins, Their Ligands, and Clinical Implications. *Neurology* **2012**, *78* (12), 914–920. https://doi.org/10.1212/WNL.0b013e31824c4728.
39. Edwards, D. R.; Handsley, M. M.; Pennington, C. J. The ADAM Metalloproteinases. *Mol. Aspects Med.* **2008**, *29* (5), 258–289. https://doi.org/10.1016/j.mam.2008.08.001.
40. Benarroch, E. E. ADAM Proteins, Their Ligands, and Clinical Implications. *Neurology* **2012**, *78* (12), 914–920. https://doi.org/10.1212/WNL.0b013e31824c4728.
41. Fukata, Y.; Chen, X.; Chiken, S.; Hirano, Y.; Yamagata, A.; Inahashi, H.; Sanbo, M.; Sano, H.; Goto, T.; Hirabayashi, M.; Kornau, H.-C.; Prüss, H.; Nambu, A.; Fukai, S.; Nicoll, R. A.; Fukata, M. LGI1–ADAM22–MAGUK Configures Transsynaptic Nanoalignment for Synaptic Transmission and Epilepsy Prevention. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (3), e2022580118. https://doi.org/10.1073/pnas.2022580118.
42. Liu, H.; Shim, A. H. R.; He, X. Structural Characterization of the Ectodomain of a Disintegrin and Metalloproteinase-22 (ADAM22), a Neural Adhesion Receptor Instead of Metalloproteinase: Insights on ADAM Function. *J. Biol. Chem.* **2009**, *284* (42), 29077–29086. https://doi.org/10.1074/jbc.M109.014258.
43. Fukata, Y.; Hirano, Y.; Miyazaki, Y.; Yokoi, N.; Fukata, M. Trans-Synaptic LGI1–ADAM22–MAGUK in AMPA and NMDA Receptor Regulation. *Neuropharmacology* **2021**, *194*, 108628. https://doi.org/10.1016/j.neuropharm.2021.108628.
44. Yamaguchi, T.; Okatsu, K.; Kubota, M.; Mitsumori, A.; Yamagata, A.; Fukata, Y.; Fukata, M.; Shibata, M.; Fukai, S. Structural Insights into Heterohexameric Assembly of Epilepsy-Related Ligand-Receptor Complex LGI1-ADAM22. *eLife* **2025**, *14*, RP105918. https://doi.org/10.7554/eLife.105918.
45. Sell, J.; Rahmati, V.; Kempfer, M.; Irani, S. R.; Ritzau-Jost, A.; Hallermann, S.; Geis, C. Comparative Effects of Domain-Specific Human Monoclonal Antibodies Against LGI1 on Neuronal Excitability. *Neurol. Neuroimmunol. Neuroinflamm.* **2023**, *10* (3), e200096. https://doi.org/10.1212/NXI.0000000000200096.
46. Goldsmith, A. P.; Gossage, S. J.; ffrench-Constant, C. ADAM23 Is a Cell-Surface Glycoprotein Expressed by Central Nervous System Neurons. *J. Neurosci. Res.* **2004**, *78* (5), 647–658. https://doi.org/10.1002/jnr.20320.
47. Michelucci, R.; Pulitano, P.; Di Bonaventura, C.; Binelli, S.; Luisi, C.; Pasini, E.; Striano, S.; Striano, P.; Coppola, G.; La Neve, A.; Giallonardo, A. T.; Mecarelli, O.; Serioli, E.; Dazzo, E.; Fanciulli, M.; Nobile, C. The Clinical Phenotype of Autosomal Dominant Lateral Temporal Lobe Epilepsy Related to Reelin Mutations. *Epilepsy Behav.* **2017**, *68*, 103–107. https://doi.org/10.1016/j.yebeh.2016.12.003.
48. Dazzo, E.; Fanciulli, M.; Serioli, E.; Minervini, G.; Pulitano, P.; Binelli, S.; Di Bonaventura, C.; Luisi, C.; Pasini, E.; Striano, S.; Striano, P.; Coppola, G.; Chiavegato, A.; Radovic, S.; Spadotto, A.; Uzzau, S.; La Neve, A.; Giallonardo, A. T.; Mecarelli, O.; Tosat-

to, S. C. E.; Ottman, R.; Michelucci, R.; Nobile, C. Heterozygous Reelin Mutations Cause Autosomal-Dominant Lateral Temporal Epilepsy. *Am. J. Hum. Genet.* **2015**, *96* (6), 992–1000. https://doi.org/10.1016/j.ajhg.2015.04.020.
49. Berghuis, B.; Brilstra, E. H.; Lindhout, D.; Baulac, S.; de Haan, G. J.; van Kempen, M. Hyperactive Behavior in a Family with Autosomal Dominant Lateral Temporal Lobe Epilepsy Caused by a Mutation in the LGI1/Epitempin Gene. *Epilepsy Behav*. **2013**, *28* (1), 41–46. https://doi.org/10.1016/j.yebeh.2013.03.032.
50. DeFelipe, J.; Fernández-Gil, M. Á.; Kastanauskaite, A.; Palacios Bote, R.; Gañán Presmanes, Y.; Ruiz, M. T. Macroanatomy and Microanatomy of the Temporal Lobe. *Semin. Ultrasound CT MRI* **2007**, *28* (6), 404–415. https://doi.org/10.1053/j.sult.2007.09.003.
51. Brown, A. S. A Review of the Déjà Vu Experience. *Psychol. Bull.* **2003**, *129* (3), 394–413. https://doi.org/10.1037/0033-2909.129.3.394.
52. Spatt, J. Déjà Vu: Possible Parahippocampal Mechanisms. *J. Neuropsychiatry Clin. Neurosci*. **2002**, *14* (1), 6–10. https://doi.org/10.1176/appi.neuropsych.14.1.6.
53. Bancaud, J.; Brunet-Bourgin, F.; Chauvel, P.; Halgren, E. Anatomical Origin of Déjà Vu and Vivid "Memories" in Human Temporal Lobe Epilepsy. *Brain* **1994**, *117* (1), 71–90. https://doi.org/10.1093/brain/117.1.71.
54. Chen, Y.; Zhu, Y.; Zhong, W.; He, J.; Gou, H.; *et al.* Identification of Autosomal Dominant Lateral Temporal Epilepsy Caused by a Novel Mutation in RELN in China: A Case Report. *Acta Epileptol*. **2024**, *6* (1), 31. https://doi.org/10.1186/s42494-024-00179-y.
55. Dazzo, E.; Leonardi, E.; Belluzzi, E.; Malacrida, S.; Vitiello, L.; Greggio, E.; Tosatto, S. C. E.; Nobile, C. Secretion-Positive LGI1 Mutations Linked to Lateral Temporal Epilepsy Impair Binding to ADAM22 and ADAM23 Receptors. *PLoS Genet.* **2016**, *12* (10), e1006376. https://doi.org/10.1371/journal.pgen.1006376.
56. Fukata, Y.; Hirano, Y.; Miyazaki, Y.; Yokoi, N.; Fukata, M. Trans-Synaptic LGI1–ADAM22–MAGUK in AMPA and NMDA Receptor Regulation. *Neuropharmacology* **2021**, *194*, 108628. https://doi.org/10.1016/j.neuropharm.2021.108628.
57. Lovero, K. L.; Fukata, Y.; Granger, A. J.; Fukata, M.; Nicoll, R. A. The LGI1–ADAM22 Protein Complex Directs Synapse Maturation through Regulation of PSD-95 Function. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (30), E4129–E4137. https://doi.org/10.1073/pnas.1511910112.
58. Braunsdorf, M.; Blazquez Freches, G. B.; Roumazeilles, L.; Eichert, N.; Schurz, M.; Uithol, S.; Bryant, K. L.; Mars, R. B. Does the Temporal Cortex Make Us Human? A Review of Structural and Functional Diversity of the Primate Temporal Lobe. *Neurosci. Biobehav. Rev*. **2021**, *131*, 400–410. https://doi.org/10.1016/j.neubiorev.2021.08.032
59. Fukata, Y.; Adesnik, H.; Iwanaga, T.; Bredt, D. S.; Nicoll, R. A.; Fukata, M. Epilepsy-Related Ligand/Receptor Complex LGI1 and ADAM22 Regulate Synaptic Transmission. *Science* **2006**, *313* (5794), 1792–1795. https://doi.org/10.1126/science.1129947.
60. Boillot, M.; Lee, C.-Y.; Allene, C.; Leguern, E.; Baulac, S.; Rouach, N. LGI1 Acts Presynaptically to Regulate Excitatory Synaptic Transmission during Early Postnatal Development. *Sci. Rep*. **2016**, *6*, 21769. https://doi.org/10.1038/srep21769.
61. Silva, J.; Wang, G.; Cowell, J. K. Temporal and Spatial Expression Pattern of the LGI1 Epilepsy Predisposition Gene during Mouse Embryonic Cranial Development. *BMC Neurosci*. **2011**, *12*, 43. https://doi.org/10.1186/1471-2202-12-43.
62. Ribeiro, P. A. O.; Sbragia, L.; Gilioli, R.; Langone, F.; Conte, F. F.; Lopes-Cendes, I. Expression Profile of Lgi1 Gene in Mouse Brain during Development. *J. Mol. Neurosci*. **2008**, *35* (3), 323–329. https://doi.org/10.1007/s12031-008-9096-0.
63. Leypoldt, F.; Wandinger, K.-P.; Bien, C. G.; Dalmau, J. Autoimmune Encephalitis. *Eur. Neurol. Rev*. **2013**, *8* (1), 31–37. https://doi.org/10.17925/ENR.2013.08.01.31.
64. Graus, F.; Titulaer, M. J.; Balu, R.; Benseler, S.; Bien, C. G.; Cellucci, T.; Cortese, I.; Dale, R. C.; Gelfand, J. M.; Geschwind, M.; Glaser, C. A.; Honnorat, J.; Höftberger, R.; Iizuka, T.; Irani, S. R.; Lancaster, E.; Leypoldt, F.; Prüss, H.; Rae-Grant, A.; Reindl, M.; Rosenfeld, M. R.; Rostásy, K.; Saiz, A.; Venkatesan, A.; Vincent, A.; Wandinger, K.-P.; Waters, P.; Dalmau, J. A Clinical Approach to Diagnosis of Autoimmune Encephalitis. *Lancet Neurol.* **2016**, *15* (4), 391–404. https://doi.org/10.1016/S1474-4422(15)00401-9.
65. Höftberger, R.; Armangue, T.; Leypoldt, F.; Graus, F.; Dalmau, J. Clinical Neuropathology Practice Guide 4-2013: Post-Herpes Simplex Encephalitis: N-Methyl-D-Aspartate Receptor Antibodies Are Part of the Problem. *Clin. Neuropathol*. **2013**, *32* (4), 251–254. https://doi.org/10.5414/NP300666.
66. Hang, H.; Zhang, J.; Chen, D.; Lu, J.; Shi, J. Clinical Characteristics of Cognitive Impairment and 1-Year Outcome in Patients with Anti-LGI1 Antibody Encephalitis. *Front. Neurol*. **2020**, *11*, 852. https://doi.org/10.3389/fneur.2020.00852.
67. Newman, M. P.; Blum, S.; Wong, R. C. W.; Scott, J. G.; Prain, K.; Wilson, R. J.; Gillis, D. Autoimmune Encephalitis. *Intern. Med. J.* **2016**, *46* (2), 148–157. https://doi.org/10.1111/imj.12974.
68. van Sonderen, A.; Petit-Pedrol, M.; Dalmau, J.; Titulaer, M. J. The Value of LGI1, Caspr2, and Voltage-Gated Potassium Channel Antibodies in Encephalitis. *Nat. Rev. Neurol.* **2017**, *13* (5), 290–301. https://doi.org/10.1038/nrneurol.2017.43.
69. Ohkawa, T.; Fukata, Y.; Yamasaki, M.; Miyazaki, T.; Yokoi, N.; Takashima, H.; Watanabe, M.; Watanabe, O.; Fukata, M. Autoantibodies to Epilepsy-Related LGI1 in Limbic Encephalitism Neutralize LGI1–ADAM22 Interaction and Reduce Synaptic AMPA Receptors. *J. Neurosci*. **2013**, *33* (46), 18161–18174. https://doi.org/10.1523/JNEUROSCI.3506-13.2013.

Author

Cara de Wesselow is a Grade 11 student with a strong interest in medical and health sciences, particularly neurobiology and molecular biology. She intends to pursue further study in medicine or biomedical sciences.

DOI: 10.36838/IJHSR819.4

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Anti-Price Gouging Laws and PPE Availability During the COVID-19 Pandemic in the United States

Garvita Jain

Dubai International Academy, Emirates Hills, First Al Khail St - Al Thanyah Third - Emirates Hills, Dubai, 00000, UAE; jain.garvita09@gmail.com
Mentor: Iva Bimpli

ABSTRACT: During the COVID-19 pandemic, severe shortages of personal protective equipment (PPE) placed significant strain on healthcare systems across the United States. At the same time, most U.S. states enacted anti-price gouging laws in response to health care emergencies, raising concerns about unintended effects on the availability of essential supplies such as PPE. This study examines the association between the enforcement of anti-price gouging laws and PPE availability in the United States in 2020. It is hypothesized that the activation of anti-price-gouging laws was associated with lower reported PPE availability, consistent with economic arguments that price constraints during sudden demand shocks reduce incentives for rapid supply expansion and reallocation. Using a quantitative observational design, the study analyzes aggregated national PPE shortage indicators and state-level cumulative PPE request data derived from publicly available reports. PPE availability patterns are examined across states based on enforcement status and relative strictness of price-gouging provisions. The findings show a complicated relationship between anti-gouging laws and PPE availability, since the raw and normalized request counts produced different trends among states. Although causality is not implied, the findings provide empirical evidence on how emergency price regulations coincided with PPE availability during the early pandemic period.

KEYWORDS: Economics, Microeconomics, COVID-19, Price Controls, Personal Protective Equipment.

Introduction

In 2020, during the COVID-19 pandemic, widespread shortages of PPE placed considerable strain on the U.S. healthcare systems.[1] For example, "87% of nurses reported having to reuse a single-use disposable mask or N95 respirator."[2] After March 18, 2020, when COVID-19 was officially declared a national emergency, anti-price gouging laws were activated by 37 states in response to the pandemic.[3] These laws restrict sharp price increases during periods of crisis to protect consumers from exploitation.[4] While this is the intention, economic theory suggests that price constraints during sudden demand shocks are likely to affect supply incentives and product availability by weakening price signals that would otherwise encourage rapid reallocation and expansion of supply.

Many standard models say that prices serve as signaling mechanisms that encourage resources to be allocated to where demand is highest. In these models, price increases may encourage producers and suppliers to expand output or redirect goods to regions with shortages. They add that when prices are constrained artificially, shortages are likely to occur due to supply failing to adjust to the boost in demand.[5] On the other hand, some argue that anti-price-gouging laws are justified on ethical grounds because sharp price increases during emergencies may undermine equity and limit access to essential goods for disadvantaged populations.[6] Besides ethical aspects, some perspectives in political economy also highlight aspects related to how markets, during crises, may be influenced because of aspects like institutional capacity, as well as issues of bargaining position regarding prices.[7,8] Notably, the majority of existing literature explores price controls in general market settings rather than in emergencies, which creates uncertainty in the topic about how such laws function under severe conditions. Previous studies have also found that price gouging regulations were related to economic/behavioral outcomes during COVID-19, like increased social contacts through search behavior in commercial spaces.[9] Other literature also looks into shortages of other essential goods during a similar time period, but fails to address the shortages of PPE specifically, despite PPE being central to healthcare and worker safety during the pandemic.[10] Therefore, this paper aims to examine the relationship between the regulation of anti-price gouging laws enforced during the initial stages (2020) of COVID-19 and the availability of PPE during the outbreak in the U.S. This paper is a contribution to the general scholarly debate around the regulation of emergencies by offering state-level findings on the link between anti-price gouging enforcement and the association of PPE with the COVID-19 demand shock. Given the limited pre-pandemic PPE availability data, this study focuses on variation in reported shortages during the early phase of the COVID-19 pandemic.

Anti-price-gouging laws are a key policy used by U.S. states in response to supply shocks, so understanding the effects of these laws during health emergencies is critical. The COVID-19 pandemic exposed severe vulnerabilities in medical supply chains, with PPE shortages directly affecting healthcare workers' safety. Data collected from July 28, 2020, indicated that hundreds of medical workers in the U.S died due to COVID-19, while many more got other sicknesses because

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.5

of it. This emphasizes the vital role of PPE in the healthcare sector by highlighting the "exposure to the virus" due to the shortage.[2] The healthcare supply chain has continued to be an area of concern throughout the COVID-19 pandemic, with hospitals experiencing issues with getting PPE, even one year into the crisis.[11] Additionally, a five-week study shared that from their sample, "46.5% of facilities lacked PPE and/or staff."[12] When well-meaning price controls unintentionally impact the supply side during times of crisis, policymakers might be forced to reconsider the nature of these laws during emergency scenarios. Furthermore, little empirical evidence exists regarding the mechanisms of anti-price-gouging laws during times of severe demand shocks.[13] Moreover, state-level differences in PPE shortage data can result from other COVID-19-related factors, like hospital density, population size, or access to procurement strategies, rather than just from the legal enforcement of price regulations. This study suggests that the enforcement of anti-price gouging laws during COVID-19 is associated with reduced short-run reported availability of PPE in 2020, measured by reported shortages and fulfillment delays. However, because the analysis is observational and descriptive, it cannot establish causality.

Methods

This study uses a quantitative descriptive research design with the use of secondary data to explore the relationship between the enforcement of the anti-price gouging laws in the U.S and reported PPE availability during the early stages of the COVID-19 pandemic. Moreover, the study combines temporal (August-September versus October-December 2020) as well as cross-sectional (state-level) comparisons for the analysis. The study is observational and analyzes existing real-world data without manipulating prices, policies, or supply conditions.

PPE availability data were primarily obtained from compilations of publicly available supply-chain reports by Get Us PPE.[14] While some of the underlying sources incorporated into the dataset originate from prior empirical or experimental studies, the present analysis uses only aggregated outcome measures and does not retain the original experimental structure. The data combines information from healthcare providers regarding these shortages, order fulfillment delays, and other unmet demand for PPE. The study analyzes data submitted between August 2020 and December 2020, which includes information on the severity of shortages and state-level information on requests and availability. This time period was selected specifically because the Get Us PPE reporting during this period provides consistent national coverage and captures the persistence of the shortages during the first year of the pandemic. Data on the enforcement of anti-price gouging laws in U.S. states were obtained from accessible state statutes and executive orders, which were supplemented by publicly available legal analyses and policy summaries.[15-17]

States were categorized depending on whether anti-price-gouging restrictions had been activated during the research period and the strictness of those laws. The strictness of the anti-price gouging laws in this study was measured qualitatively through several factors, including low price increases, inclusion of basic commodities, severity of penalties, and enforcement intensity, among others. States were characterized as "strict" when the laws contained several restrictive provisions simultaneously, including low permissible price increases, wide product coverage, and large penalties, while states with fewer of these characteristics were termed as relatively less strict. The data on these factors were also derived from state statutes and summarized in publicly available legal analyses.[18-20] Although such a classification is not a numerical index for strictness, certain comparisons can still be made between the more and less restrictive states.

PPE availability was proxied using quantitative indicators derived from the Get Us PPE dataset, including the number of reported PPE shortage incidents, reported fulfillment delays, and the frequency of unmet PPE requests. Since there was variability in PPE requests between states, the cumulative PPE requests intensity was also categorized so that there is a better comparison between states with different enforcement intensities. The primary independent variable was the presence of anti-price gouging enforcement, operationalized as a binary indicator (1 = enforcement active following an emergency declaration; 0 = no active enforcement).

The analytical framework treats anti-price gouging enforcement as the independent variable and PPE availability indicators as dependent variables. It treats price gouging enforcement as a policy constraint on prices, with the intensity of PPE request serving as a proxy measure of unmet demand. The analysis will be descriptive, focusing on correlating differences in PPE shortages across states with price-gouging enforcement and states without such enforcement. In order to assess the differences in the request intensity of PPE, the states were grouped based on ranges of PPE requests: 0-99, 100-199, 200-299, 300-399, 400-499, and 500+. Variation over time following the COVID-19 emergency declaration in the U.S was used to document the national severity and persistence of PPE shortages during the study period. In addition, the strictness of anti-price gouging laws was used to examine whether more restrictive enforcement states were associated with higher reported shortage intensity compared to more flexible regulatory structures. Therefore, the analysis shows associations rather than causality.

Data for PPE requests and shortage data are subject to reporting limitations. This is due to submissions to public reporting platforms being voluntary and uneven across time periods and states as well. Therefore, the dataset is likely to underestimate the true demand during peak periods because the dataset might not take into account all PPE shortages and other reported figures. In addition, even the reporting intensity itself may be related to the size of the healthcare systems in the states, meaning that more requests may be due to better reporting networks as opposed to a shortage of PPE. Given this variation, the analysis relies on reported PPE shortage indicators and percentage-based measures rather than population-normalized rates. Because of this, the analysis looks into patterns and associations rather than exact levels of PPE availability.

To address the potential influence of state size on raw request counts, PPE requests were also normalized by dividing each state's cumulative request count by its estimated number of healthcare workers (in thousands), as reported by the U.S. Bureau of Labor Statistics Occupational Employment Statistics (2020), and multiplying by 100 to yield a rate of PPE requests per 100,000 healthcare workers. This normalization accounts for the fact that larger states with larger healthcare workforces will generate more absolute requests regardless of shortage severity, and allows for a more equitable cross-state comparison of request intensity.

In addition to the descriptive analysis, an Ordinary Least Squares (OLS) regression was estimated to examine the association between enforcement status and normalized PPE request intensity while controlling for key confounders. The dependent variable was PPE requests per 100,000 healthcare workers. The independent variable of interest was enforcement status, operationalized as a binary dummy variable (1 = active enforcement; 0 = no active enforcement). Control variables were included to account for three potential confounders identified by the study's design limitations: (1) COVID-19 cumulative case rate per 100,000 population through December 2020, obtained from the Centers for Disease Control and Prevention,21 to control for pandemic severity; (2) the natural log of state population (U.S. Census 2020),22 to control for state size; and (3) healthcare worker density, defined as healthcare workers as a percentage of state population, to control for the scale of each state's health services infrastructure. This model does not demonstrate causality since the observational design prevents causal inference, but it does allow for a more rigorous descriptive assessment of the enforcement-shortage association while holding key confounders constant.

Results and Discussion

This section presents descriptive and correlational results on PPE availability during 2020.

The analysis focused on the indicators of the availability of PPE during the period when the enforcement of anti-price-gouging policies was active among US states. The data were collected in aggregate form and were submitted between August and December 2020. The data were collected at the state level where applicable. Additionally, the data highlight the trends of PPE shortages as well as the unmet requests over time. It also shows the comparison between states that had active enforcement of anti-price gouging laws and those that did not.

National PPE Shortage Trends (August-December 2020):

Before examining policy-related differences across states, national PPE shortage trends are summarized in Table 1 to establish the broader supply environment during the study period.

Table 1: National PPE shortage indicators reported in the United States, August–December 2020.

Time period	Reported % facilities with no PPE supply	Cumulative PPE requests (March-month)	Reported % unmet PPE requests (cumulative)	Reporting months
Aug-Sep 2020	77%-80%	16,404 - 16,654	~88%	August, September
Oct-Dec 2020	63%-70%	16,808 - 19,981	>85%-88%	October, November, December

Table 1 presents the cumulative PPE shortage indicators that were reported by Get Us PPE between August and December of 2020. Throughout this study period, a clearly high proportion of reporting facilities indicated having no available PPE supply. However, the reported percentages did decline modestly over time. Even with this decline, the cumulative unmet PPE request rates remained above 85% throughout the study period. The total cumulative PPE requests increased across reporting months, therefore reflecting sustained demand during the pandemic. These indicators collectively suggest that PPE scarcity remained an issue in late 2020 as well, which provides a baseline context to interpret variations in request intensity between states.

PPE Request Intensity by Enforcement Status:

Table 2: Distribution of U.S. states by cumulative PPE request category and anti-price gouging enforcement status.

PPE Request Category (Cumulative)	Enforcement States (1)	Non-Enforcement States (0)
0-99	6	7
100-199	10	2
200-299	5	1
300-399	3	1
400-499	3	1
500+	10	1
Total states	37	13

Table 2 shows that enforcement states are overrepresented in the higher PPE request categories. Enforcement states have 10 states among the 500+ requests per state category, while only one non-enforcement state is represented in that same category. Furthermore, non-enforcement states are represented almost entirely within the 0-99 requests per state category. The size of the enforcement states' healthcare workforce is substantially larger than that of non-enforcement states, however. Thus, it is likely that this raw distribution of PPE requests in Table 2 reflects a bias towards overestimating the PPE shortage in enforcement states relative to non-enforcement states.

To address this, Table 3 presents the distribution of states when the number of PPE requests is normalized by the size of each state's healthcare workforce. The mean value for PPE requests per 100,000 healthcare workers is calculated using 2020 Bureau of Labor Statistics data on healthcare employment.[23]

DOI: 10.36838/IJHSR819.5

Table 3: Distribution of U.S. states by normalized PPE request intensity (per 100,000 healthcare workers) and enforcement status.

Requests per 100,000 Healthcare Workers	Enforcement States (1)	Non-Enforcement States (0)
0-49	0	2
50-99	7	2
100-149	20	4
150-199	10	0
200+	0	5
Mean (SD)	**129.8 (36.9)**	**339.9 (423.5)**

The results of this calculation reveal that the distribution of request rates per 100,000 healthcare workers differs from that in Table 2. Most notably, the mean request rate for non-enforcement states is 339.9 per 100,000 healthcare workers, compared with 129.8 per 100,000 healthcare workers for enforcement states. Furthermore, a two-sample t-test reveals that these differences in mean request rates are statistically significant ($t = -3.04$, $p = 0.004$). These results help explain the differences in Table 2: the non-enforcement states with the highest PPE request rates include small states like Rhode Island, Delaware, and New Hampshire, each of which has a relatively low healthcare workforce by state standards.

These results are important for recognizing the confound between population and healthcare workforce sizes in the raw data presented in Table 2. Thus, PPE requests per 100,000 healthcare workers are a more accurate data point for enforcement and non-enforcement states than raw PPE requests per state.

To determine whether the association between enforcement status and PPE request intensity remains after adjusting for key confounders, an ordinary least squares (OLS) regression was performed with PPE requests per 100,000 healthcare workers as the dependent variable. The results are presented in Table 4.

Table 4: OLS regression results: PPE requests per 100,000 healthcare workers on enforcement status and controls (N = 50 states).

Variable	Coefficient	Std. Error	t-statistic	p-value
Intercept	−23.25	554.75	−0.04	0.967
Enforcement (dummy)	−269.00	116.44	−2.31	0.026*
COVID cases per 1,000/100k	−10.67	34.93	−0.31	0.761
Log(Population)	39.45	48.64	0.81	0.422
Healthcare worker density (%)	36.06	116.09	0.31	0.758
R^2	**0.183**			
Adjusted R^2	**0.110**			

The enforcement coefficient is negative and statistically significant ($\beta = -269.00$, $p = 0.026$). Controlling for pandemic severity, state population, and healthcare personnel density, there were 269 fewer PPE requests per 100,000 healthcare personnel in enforcement states than in non-enforcement states. This result is consistent with the normalized descriptive findings in Table 3 and implies that the raw trend in Table 2 was mainly caused by the variation in the size of the states.

There is moderate explanatory power of the regression model (adjusted $R^2 = 0.110$) with all control variables being nonsignificant. This result is expected given a small sample ($n = 50$) and high diversity across states within the country. The coefficient on the COVID case severity variable is statistically insignificant, meaning that the difference in normalized requests cannot be explained solely by pandemic severity. It is important to consider these conclusions cautiously as they may be affected by several factors. In particular, an observational study design allows making inferences about possible unobserved factors that might be influential in the relationship in question. These include, for instance, state-level purchasing capability, density of hospitals, and reporting structure.

Accordingly, these findings do not establish a causal relationship between anti-price gouging enforcement and PPE availability, but offer a more methodologically rigorous descriptive assessment than the raw request counts alone.

Variation in PPE Requests by Strictness of Anti-Price Gouging Laws:

Within enforcement states, PPE request intensity varied by the relative strictness of anti-price-gouging laws. Among the 7 states that were classified as having stricter regulations, 4 of these states fell into the 500+ category, whereas 1 fell into the 400-499 category. This suggests that stricter regulatory environments were more frequently associated with the highest levels of reported PPE request intensity. Therefore, states classified as having stricter regulations were more frequently represented in higher cumulative PPE request categories, including the 300+ and 500+ groups. These differences are descriptive and reflect observed variation within the enforcement group.

Beyond the differences between active enforcement and non-enforcement states, additional variation in PPE request intensity could be seen within the active enforcement states, primarily due to the design of anti-price gouging provisions. States that were classified as having stricter regulations, such as low price increase thresholds or substantial penalties, were more frequently represented in higher cumulative PPE request categories than those with more flexible laws and standards. While the noted patterns are descriptive, they suggest that, rather than the enforcement status alone, legal design features could also be a factor that is related to differences in reported supply constraints in response to extreme demand shocks. This is an important observation because it implies that regulatory intensity may shape the degree to which price adjustments are constrained during emergencies.

These patterns, however, may also reflect other state-level differences like COVID-19 case severity, hospital density, population size, procurement strategies, and reporting intensity rather than the laws themselves.

Interpretation Through Economic Theory:

The findings of the study indicate a clear association between the enforcement of anti-price gouging laws and reported PPE shortages during 2020, during the COVID-19 pandemic in the U.S. Specifically, states with active enforcement were disproportionately represented in higher cumulative PPE request categories compared to states without active enforcement. It could be seen in the results that states that had active enforcement of these laws had a higher number in the higher cumulative PPE request categories compared to the states with no active enforcement. These specific state-level patterns occurred within a broader and pervasive PPE shortage nationwide, which suggests that the differences across the policy environments come under the sustained supply constraints as a whole rather than isolated disruptions. Even a year into the COVID-19 pandemic, supply chain shortages and demand surges for PPE continued to be a problem.[11] Moreover, national surveys, like one conducted by the APIC, confirmed PPE supply shortages in the U.S. during the COVID-19 pandemic, supporting that the high levels of unmet demand observed in this study were not just due to reporting.[24]

The observed correlation is consistent with key economic frameworks that emphasize the role of prices as signals for allocating scarce resources during times of sudden demand shocks. When prices are constrained, suppliers would be less incentivized in the short run to either expand production, enter markets, or quickly reallocate resources towards specific goods and services that are facing shortages. In the context of the pandemic, where PPE demand rose sharply, and supply chains faced significant production constraints and logistical disruptions, restrictions on price adjustments may have contributed to shortages, which were expressed through unmet requests rather than higher prices.[25,26] From this perspective, anti-price-gouging laws may reduce the effectiveness of market coordination during emergency resource distribution as they reduce the potential for profits to incentivize rapid scaling and allocation. This mechanism is, however, only theoretical within the context of this study, as the descriptive method does not allow for direct measurement.

Ethical and Political Economy Considerations:

At the same time, these findings should not be interpreted as a normative critique of anti-price-gouging laws. Such regulations are usually justified on ethical grounds related to issues of fairness and equity, especially in situations where emergency goods are being supplied. The fear that price gouging could deny vulnerable groups access to certain goods has been a key factor in justifying price gouging restrictions.[6,27] Additionally, some researchers believe that emergency markets could also be subject to panic buying, information asymmetries, and bargaining power imbalances, suggesting the price increases observed might not necessarily lead to efficiency and equilibrium.[28] This implies that, in such situations, alternative non-price systems to ensure the availability of critical products during a crisis may be necessary.

Alternative Explanations and Study Limitations:

In this study, several limitations should be acknowledged. Shortage and request data for PPE were collected from voluntary reporting platforms, which might not reflect actual shortage and request numbers and might have some inconsistencies in terms of reporting patterns across states and time. There might have been differences in population, healthcare facilities, pandemic spread, and timings of outbreaks, which could have also affected demand for PPE, irrespective of the price-gouging policies. These factors suggest that enforcement status may be correlated with underlying structural differences that also influence PPE availability, making it difficult to isolate the role of anti-price-gouging laws. Furthermore, considering the fact that the study is based on an observational cross-sectional design and does not involve randomization or pre-interventional baselines, it cannot support causal claims even if statistical tests were applied. Associations may result from hidden confounders, without a need for an actual causative role played by anti-price-gouging intervention measures. Therefore, the results should be interpreted as indicative associations rather than definitive estimates of policy effects. In the current study, partial correction for these confounders has been achieved by means of standardizing with respect to the number of healthcare workers and OLS regression analysis, controlling for the severity of COVID-19 cases, state population, and healthcare workers' density. At the same time, some additional variables, such as the development of procurement systems at the state level, political economy, and voluntary reporting, have yet to be considered. Further research could focus on panel regressions and differences in differences using variations within states and between times.

Implications and Directions for Future Research:

Despite these limitations, this study can be seen as a contribution to empirical literature as it offers a descriptive relationship between emergency enforcement of price gouging and the availability of PPE during an unprecedented public health crisis. It stresses the need to examine how regulatory interventions may interact with supply-side incentives and distribution constraints during emergencies, not just theoretically or normatively, but also empirically. These findings can be used as a guide for future evaluations of emergency price regulations by emphasizing the value of taking into account supply-side considerations as well as consumer protection objectives. More broadly, the research contributes to the policy debate by suggesting that emergency market interventions may need to be accompanied by additional measures, including coordination of procurement efforts, the provision of rapid production scaling subsidies, or increases in strategic stocks, in order to alleviate shortages and protect consumers. The study also shows the value of making use of real-world data on shortages and requests as a way of evaluating emergency policies, which provides a foundation for future research aimed at finding out mechanisms by which emergency policies interact with the supply chain and impact access to essential medical products like PPE.

DOI: 10.36838/IJHSR819.5

Conclusion

This study aimed to explore the association between anti-price gouging enforcement and PPE supply, as indicated by PPE availability reports, during the early phase of the COVID-19 pandemic in the United States. Using aggregated U.S. shortage indicators and state-level cumulative PPE requests, the study found that although states implementing an active policy of price-gouging enforcement appeared to be more likely to be found among the high raw PPE requests, the opposite situation occurred when taking into account the number of healthcare personnel. States with an enforcement policy demonstrated significantly fewer requests for personal protective equipment per 100,000 healthcare personnel compared to states without any enforcement policies (129.8 vs. 339.9; p=0.004). This effect was confirmed by the results of OLS regression analysis. Moreover, there was evidence of differences within states with anti-price-gouging enforcement itself, showing a relation to the legal design of the policies classified by their strictness, implying that the structure of price-gouging provisions may also be relevant to reported supply outcomes.

While the study did not establish causal effects, the outcomes contribute to the literature by providing descriptive, empirical evidence on the association between emergency price regulation and PPE supply during a large-scale health crisis. By focusing on state-level variation during an extreme demand shock, this analysis extends the existing debates over price controls beyond purely theoretical and opinion-based discussions. It points out the need to understand the significance of assessing emergency market interventions in real-life contexts, where ethical goals such as consumer welfare need to be traded off with the need to ensure that adequate supplies of critical goods are maintained. This may be of use to policymakers and researchers dealing with future health crises, where they may benefit from combining price controls with other supply-side policies and monitoring of critical health resources.

Acknowledgments

The author thanks Dr. Iva Bimpli and Daryn Howland for guidance and feedback during the development of this study.

References

1. CDC. *Conserving Supplies of Personal Protective Equipment in Healthcare Facilities during Shortages*. Healthcare Workers. https://www.cdc.gov/niosh/healthcare/hcp/pandemic/conserving-PPE-shortages.html (accessed 2026-01-02).
2. Cohen, J.; Rodgers, Y. van der M. Contributing Factors to Personal Protective Equipment Shortages during the COVID-19 Pandemic. *Prev. Med.* **2020**, *141*, 106263. https://doi.org/10.1016/j.ypmed.2020.106263.
3. *Anti–Price Gouging Laws: Why a Pandemic Is Different from Other "Emergencies."* Cato Institute. https://www.cato.org/regulation/summer-2022/anti-price-gouging-laws-why-pandemic-different-other-emergencies (accessed 2026-01-02).
4. *Anti-Price-Gouging Laws Entrench Shortages*. Cato Institute. https://www.cato.org/commentary/anti-price-gouging-laws-entrench-shortages (accessed 2026-01-02).
5. *The Use of Knowledge in Society.* Econlib. https://www.econlib.org/library/Essays/hykKnw.html (accessed 2026-01-03).
6. Zwolinski, M. The Ethics of Price Gouging. *Bus. Ethics Q.* **2008**, *18* (3), 347–378. https://doi.org/10.5840/beq200818322.
7. Christopher, M.; Peck, H. Building the Resilient Supply Chain. *Int. J. Logist. Manag.* **2004**, *15*, 1–13. https://doi.org/10.1108/09574090410700275.
8. Rodrik, D. Institutions for High-Quality Growth: What They Are and How to Acquire Them. NBER Working Paper No. 7540; National Bureau of Economic Research: Cambridge, MA, 2000. https://doi.org/10.3386/w7540
9. Chakraborti, R.; Roberts, G. How Price-Gouging Regulation Undermined Covid-19 Mitigation: County-Level Evidence of Unintended Consequences. *Public Choice* **2023**, 1–33. https://doi.org/10.1007/s11127-023-01054-z (accessed 2026-02-01).
10. *(PDF) Anti-Price Gouging Laws, Shortages, and COVID-19: Insights from Consumer Searches.* https://www.researchgate.net/publication/342092149_Anti-Price_Gouging_Laws_Shortages_and_COVID-19_Insights_from_Consumer_Searches (accessed 2026-02-01).
11. *Premier Data: The State of PPE Supply One Year into COVID-19.* Premier. https://premierinc.com/newsroom/blog/premier-data-the-state-of-ppe-supply-one-year-in-to-covid-19 (accessed 2026-02-01).
12. Gibson, D. M.; Greene, J. State Actions and Shortages of Personal Protective Equipment and Staff in U.S. Nursing Homes. *J. Am. Geriatr. Soc.* **2020**, *68* (12), 2721–2726. https://doi.org/10.1111/jgs.16883 (accessed 2026-02-01).
13. (PDF) Anti-Price Gouging Laws, Shortages, and COVID-19: Insights from Consumer Searches. *ResearchGate* **2025**. https://doi.org/10.2139/ssrn.3613726.
14. *PPE Shortage Data | real-time metrics from Get Us PPE*. Get Us PPE. https://getusppe.org/data/ (accessed 2026-01-18).
15. *Corporate Compliance, Professional Perspective – Price Gouging Laws and COVID-19: What Supply Chain Businesses Should Know.* https://www.bloomberglaw.com/external/document/XC7FNJKC000000/corporate-compliance-professional-perspective-price-gouging-laws (accessed 2026-02-01).
16. *Price Gouging State Statutes.* https://www.ncsl.org/financial-services/price-gouging-state-statutes (accessed 2026-02-01).
17. Microsoft Word - General Reference Chart Re US Price Gouging Laws. DOCX. https://www.mayerbrown.com/-/media/files/perspectives-events/publications/2020/04/general-reference-chart-re-us-price-gouging-laws.pdf (accessed 2026-02-01).
18. Osborn, K. L.; Mahoney, J.; Behrmann, A. M.; Jacobson, J. S. *Price Gouging Laws and COVID-19: Navigating Unprecedented Market Conditions | Faegre Drinker Biddle & Reath LLP*. https://www.faegredrinker.com/en/insights/publications/2020/3/price-gouging-laws-and-covid-19 (accessed 2026-02-01).
19. *COVID-19: State Attorneys General Enforcement of Price Gouging Laws*. https://www.orrick.com/en/insights/2020/03/covid-19-state-attorneys-general-enforcement-of-price-gouging-laws (accessed 2026-02-01).
20. *Price Gouging Concerns Rise During the COVID-19 Pandemic*. Rogers Joseph O'Donnell. https://www.rjo.com/publications/price-gouging-concerns-rise-during-covid-19-pandemic/ (accessed 2026-02-01).
21. Centers for Disease Control and Prevention. COVID Data Tracker: United States COVID-19 Cases and Deaths by State. https://covid.cdc.gov/covid-data-tracker (accessed 2026-05-01).
22. U.S. Census Bureau. 2020 Census Apportionment Results: Resident Population by State. https://www.census.gov/data/tables/2020/dec/2020-apportionment-data.html (accessed 2026-05-01).

23. U.S. Bureau of Labor Statistics. Occupational Employment and Wage Statistics, State Cross-Industry Estimates, May 2020. https://www.bls.gov/oes/2020/may/oessrcst.htm (accessed 2026-05-01).
24. *National survey shows dire shortages of PPE, hand sanitizer across the U.S. - APIC*. https://apic.org/news/national-survey-shows-dire-shortages-of-ppe-hand-sanitizer-across-the-u-s/ (accessed 2026-02-02).
25. Moutii, M. *Anti-Price Gouging Laws are doing more Harm than Good*. IREF Europe EN. https://en.irefeurope.org/publications/online-articles/article/anti-price-gouging-laws-are-doing-more-harm-than-good/ (accessed 2026-02-02).
26. Chakraborti, R.; Roberts, G. Anti-Price Gouging Laws, Shortages, and COVID-19: Insights from Consumer Searches. *J. Priv. Enterp.* **2020**, *35* (Winter 20), 1–20. https://doi.org/10.2139/ssrn.3613726.
27. *The Morality of Price Gouging - The Edge, A Leader's Magazine*. https://theedgeleaders.com/the-morality-of-price-gouging/ (accessed 2026-02-01).
28. Helveston, M. N. Regulating Economic Opportunism in Postdisaster Markets. N.C. L. Rev. 2024, 102 (3), 811. https://scholarship.law.unc.edu/nclr/vol102/iss3/4.

Author

Garvita Jain is a Year 12 student (Class of 2027) at Dubai International Academy, Emirates Hills. She intends to pursue a degree in Economics or Developmental Economics. Her academic interests include market regulation, public policy, and behavioral economics.

DOI: 10.36838/IJHSR819.5

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Parkinson's Disease: Genetic Factors, Emerging Therapies, and Their Implication for the Future

Iris S. Lee

Mountain View High School, 3535 Truman Ave, Mountain View, CA 94040, USA; ii0710lee@gmail.com

ABSTRACT: Parkinson's disease is a progressive neurodegenerative disease where the dopaminergic neurons in the substantia nigra become damaged and die over time. Every six minutes, an American is diagnosed with Parkinson's disease, making it the fastest-growing neurological condition globally.1 Despite its scale, there has been no solution to cure this disease, truly making it a public health crisis. In this review, we will discuss gene therapy (GBA-1, SNCA, LRRK2), deep brain stimulation (adaptive, connective), and cell therapy (embryonic stem cells, induced pluripotent stem cells). Gene therapy is the process of introducing genetic material into the brain to target specific molecular pathways that have been damaged by PD. Deep brain stimulation is the implantation of electrodes in specific regions of the brain to deliver electrical pulses, helping to regulate abnormal brain signals often found in PD that cause motor symptoms. Cell therapy is the replacement of damaged neurons and restoration of dopaminergic function, decreasing the severity of motor symptoms. While these therapies are not completely developed, they have the potential to improve PD patients' quality of life by personalization, effectively slowing symptom progression and decreasing side effects. Moving forward, researchers and clinicians should respectively focus on addressing therapy concerns, such as the irreversibility of gene therapy, and closely follow these emerging therapies in order to offer their patients the best possible treatments.

KEYWORDS: Translational Medical Sciences, Disease Treatment and Therapies, Neurodegenerative Disease, Parkinson's Disease, Gene Therapy, Deep Brain Stimulation, Cell Therapy.

Introduction

Every six minutes, an American is diagnosed with Parkinson's disease, making it the fastest-growing neurological condition globally. Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by the degeneration of dopaminergic neurons in the substantia nigra, a group of nuclei part of the basal ganglia that regulates voluntary motor control.[2] In a healthy individual, dopamine removes inhibitory signals in the basal ganglia, allowing for voluntary movement. However, in PD patients, the gradual loss of dopaminergic neurons over time results in insufficient levels of dopamine, ultimately leading to bradykinesia and a struggle to initiate voluntary movements. While the cause of Parkinson's disease is still unknown, one cause that is currently the subject of intensive research is Lewy bodies, intraneuronal proteinaceous inclusions which are primarily made up of alpha-synuclein, in the neurons of the substantia nigra, disrupting nerve transmission and ultimately damaging nerves and even leading to cell death.[2,3]

The most common characteristics individuals with Parkinson's disease suffer from are motor symptoms, collectively called Parkinsonism. Parkinsonism can be caused by a variety of factors, such as drugs, toxins, and other diseases, which must be excluded before diagnosing a patient with PD. Parkinsonism motor symptoms include hand tremors, slow movements, limb rigidity, and problems with gait and balance. Additional motor symptoms often found in PD patients but not associated with Parkinsonism are slurred speech and reduced facial expressions. While the most significant symptoms of PD are motor-related, other non-motor-related symptoms include behavior and mood changes, cognitive impairment, and sleep disturbances.[2] While novel therapies are not completely developed, they have the potential to improve PD patients' quality of life by personalization, effectively slowing symptom progression and decreasing side effects. Moving forward, researchers and clinicians should respectively focus on addressing therapy concerns, such as the irreversibility of gene therapy, and closely follow these emerging therapies in order to offer their patients the best possible treatments.

Genes associated with Parkinson's Disease

1.1. SNCA:

As previously stated, the cause of PD is unknown. However, through ongoing research, some potential causes have emerged. The most significant potential cause that this research paper will explore is genetics. The SNCA gene, the first gene discovered to be related to PD, controls the production of alpha-synuclein.[4,5] Alpha-synuclein is most abundant in the brain, specifically at the presynaptic terminals of neurons.[4] In 1997, researcher Roger Duvoisin from the Parkinson's Foundation first discovered this gene to be related to PD.[5] More than 30 SNCA variants have been identified. These variants are inherited in an autosomal-dominant pattern, meaning only one SNCA gene variant is needed to significantly increase the risk of developing PD. SNCA variants also tend to cause young-onset Parkinson's disease (YOPD), which is PD occurring before

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.6

age 50. Disease progression may also be faster in genetically related cases. While research is still ongoing, researchers have discovered that SNCA variants can lead to the accumulation of Lewy bodies through two mechanisms.[5] First, the alpha-synuclein proteins do not fold or clump together properly.[4] This is due to the amino acid alanine being replaced with the amino acid threonine or proline, ultimately leading to an incorrect 3-dimensional shape.[4] Second, the variants cause an overproduction of alpha-synuclein proteins through excess copies in gene replication that cannot be destroyed fast enough, leading to clumping.[4,5]

1.2. GBA:

The GBA or GBA-1 gene is the most common PD-related gene.[5] The GBA gene codes for the lysosomal enzyme glucocerebrosidase (GCase), which maintains glycosphingolipid homeostasis.[3] Variants of this gene are found in 5-15% of people with PD.[3] Compared to other primary gene mutations of PD, variants of the GBA are considered only to increase the chances of an individual getting PD, not necessarily being a definitive cause. The majority of carriers of the GBA variant, even those with a family history of PD with the GBA variant, will not develop PD. Other risk factors in addition to the GBA variant, such as age or environment, are needed to significantly increase the risk of developing PD. Carrying a single heterozygous GBA increases the risk of developing PD, but having both genes, a homozygous variant, further increases the risk. Researchers are currently exploring the specifics of how GBA variants increase PD risk in brain cells.[3] They hypothesize that GBA variants cause a loss of GCase activity and lysosomal dysfunction, impairing alpha-synuclein metabolism and leading to aberrant lipid metabolism and an accumulation of waste.[3] Currently, pharmaceutical companies are testing drugs through clinical trials to target GBA to slow or even stop PD progression.[5]

1.3. LRRK2:

The LRRK2 gene is the second most common PD-related gene.[5] It is a large protein, about 2527 amino acids.[6] Its variants are found in 1% of people with PD and 5% of people with a genetic history.[5] Although LRRK2 genes are generally considered causative of PD, some people who carry these variants will not develop PD. Similar to some other variants, the risk of developing PD increases with age and has a dominant inheritance pattern. Unlike other gene variants, certain ethnic groups have a higher prevalence of LRRK2 variants, such as Ashkenazi Jews (about 15%) and North African Berbers (about 40%). People with the LRRK2 variant-caused PD may experience less frequent motor difficulties and slower disease progression compared to other variants. Ongoing research explores how LRRK2 specifically leads to PD.[5] Scientists believe mutations in the LRRK2 gene may cause an accumulation of α-synuclein, which is one of the main stimulants of microglial activation.[6] This activation is also believed to contribute to neuroinflammation and neuronal death in PD.[6]

Gene Therapy

2.1. Dopamine-Related Gene Therapy:

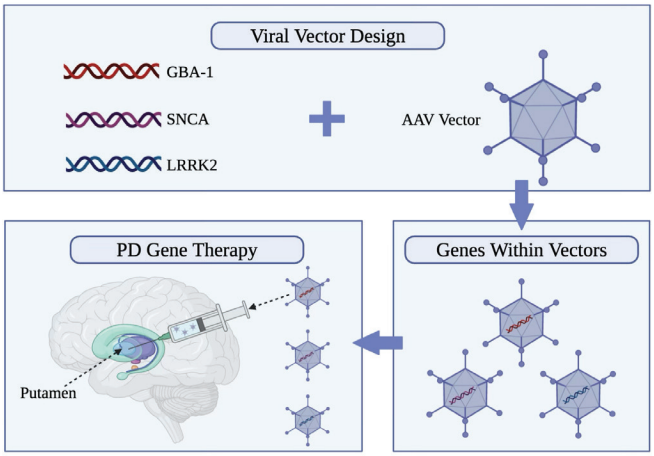


Figure 1: Adeno-associated viral (AAV) vectors carry PD-associated genes: GBA-1, SNCA, and LRRK2, for targeted delivery. The loaded AAV vectors are injected into the putamen, a motor-functioning area of the brain, to improve motor control for patients with Parkinson's disease. (Created with BioRender).

Gene therapy in the context of Parkinson's disease is the process of introducing genetic material into the brain to target specific molecular pathways that have been damaged by PD.[7] It is an emerging and promising approach to promoting dopaminergic growth and reducing neurodegeneration. It addresses both symptomatic relief and disease modification. Gene therapy will typically use viral vectors to deliver therapeutic genes directly to affected brain regions **(Figure 1)**. One main approach in PD gene therapy is Dopamine Synthesis Enhancement, which converts L-DOPA to dopamine more efficiently, addressing a primary symptom of PD: loss of dopaminergic neurons. Aromatic L-amino acid decarboxylase (AADC), an enzyme part of the dopamine synthetic apparatus, is used in Dopamine Synthesis Enhancement by being delivered to the striatum or putamen (motor-functioning areas of the brain). Human trials showed reduced "off" time, lower need for L-DOPA doses, and improved motor function specifically in early-phase trials. L-DOPA is the primary clinical approach for PD treatment because of its ability to cross the brain barrier, thus enabling oral treatment and making it easier for PD patients to take. While L-DOPA can address the loss of dopaminergic neurons in PD patients, there are some drawbacks. First, L-DOPA is only a symptomatic treatment, meaning it cannot stop disease progression, thus limiting its scope of effectiveness. Additionally, severe side effects of L-DOPA include "OFF-states": periods of immobility and incapacity often followed by depression or hypomania due to inconsistent function of the drug. These "OFF-states" have been attributed to unpredictable gastric emptying, an overload of the blood-brain barrier carrier system which transports L-DOPA, and insufficient aromatic L-amino acid decarboxylase (AADC) levels. However, new formulations of L-DOPA have improved drug absorption. Many rodent studies found striatal vector-mediated AADC overexpression to be well tolerated and cause symptom improvement. Several non-human primate studies support this, showing that once a certain

level of transfection was reached, there was a decreased need for medication and improvement of motor control when off medication. Additionally, human trials (AAV2-AADC in the putamen of severe PD patients) showed safe results, an improvement in motor symptoms, and reduced UPDRS scores. Ongoing studies now aim to optimize the delivery and dosage of L-DOPA.

2.2. Neuroprotection and Growth Factor Approaches:

Furthermore, more research is being conducted on growth factors as a treatment target for neurodegenerative diseases such as PD. The primary focus has been on the glial cell-line derived neurotrophic factor family of ligands (GFL) that encompasses glial cell-line derived neurotrophic factor (GDNF), neurturin (NRTN), artemin (ARTN), and persephin (PSPN). The GFL signals via activation of the RET receptor complex and cofactor GFRα, initiating MAP kinase and PI3-kinase pathways, which support neuron survival and neuritogenesis. Several studies have shown that direct delivery of GDNF into the SN and the striatum reduces Parkinsonian behavior and improves neuritogenesis in toxin models of PD and phase 1 trials. However, phase 2 trials have failed due to poor drug distribution and an insufficient number of residual neurons for GDNF to affect. Due to phase 2 failure, other gene therapies are being explored, such as AAV-GDNF, which was successfully injected into the caudate nucleus. Similar to L-DOPA delivery, a new approach of AAV-GDNF vectors in the putamen is safe and effective at treating Parkinsonian behavior in murine PD models.[7]

2.3. Genetic Mutation-Targeted Therapies:

2.3.1. GBA:

Another primary approach to correcting genetic mutations, specifically GBA, is GCase enzyme replacement therapy.8 While its recombinant form was not initially able to be used for PD due to its inability to cross the BBB, viral vectors of GBA are now being directly injected into the midbrain. Several mouse studies have shown that this restores GCase activity, thus reducing waste, neuroinflammation, gliosis, and preventing cognitive and motor deficits. Different virus types, such as AAV9, AAV2/5, AAV1, and promoters, had the best results. Due to success in animal trials, the PROPEL study, a 2020 phase 1/2a clinical trial sponsored by Prevail Therapeutics, is testing AAV9-GBA1 (PR001/LY3884961) in individuals with moderate to severe PD who carry at least one GBA1 mutation. This study enrolled 20 patients who were separated into low and high dose groups (exact dose not disclosed) where PR001 was administered only once through intracisternal injection in cisterna magna.[8] Close monitoring, blood and urine tests will start from Day 2 of administration to month 18, then a follow-up from month 18 to month 60.[9] This study is currently ongoing and recruiting but results are said to be released in 2029.[8,10]

2.3.2. SNCA:

Another upcoming experiment relies on the development of CRISPR-based technologies, specifically dCas9, to decrease SNCA gene expression, leading to lower alpha-synuclein protein levels.[11] By lowering alpha-synuclein protein levels, this form of gene therapy may be able to slow or halt PD progression. Scientists will use CRISPR to alter precise locations of the genome and penetrate the neuronal cells derived from induced pluripotent stem cells, which this research paper will dive into further in cell therapy. In the future, the outcomes of these pre-clinical models will be observed for their long-term effects. If the outcome of these preclinical models is favorable, then the Michael J Fox Foundation will plan on performing clinical trials and ultimately seeking regulatory approval.[11]

2.3.3. LRRK2:

The last main new approach to gene therapy targets the LRRK2 gene. A company called Denali Therapeutics, based in South San Francisco, partnered with Biogen to create drugs in pill form called DNL201 and DNL151, also known as BII122. These are kinase inhibitors, in hopes of delaying progression of PD in early-stage patients with the LRRK2 gene mutation by reducing the activity of LRRK2 proteins, thus improving the function of lysosomes.[12] In a press release from January 2020, Denali Therapeutics announced that both compounds have shown little to no side effects.[13] Later that year, in August, it was announced that while both compounds can advance to clinical trials, DNL151 is preferred as it has more flexible dosing regimens. Currently, one of the ongoing and seemingly successful Phase 2 clinical studies is LUMA, which started in April 2022 with study completion estimated in December 2025. This study is large, global, and placebo-controlled in the hopes of decreasing PD symptom progression. It enrolled 650 patients with early-stage PD aged 30 - 80 who were diagnosed with PD within 2 years of starting the study, and people with LRRK2 gene mutation and those with idiopathic PD.[13]

2.4. Gene Therapy, Conclusion:

Overall, gene therapy offers a promising symptomatic solution for PD. While mainly being a symptomatic treatment, it has the potential to improve the quality of life of PD patients with a variety of genetic mutations, with decreased progression of symptoms and increased dopamine production, meaning more motor control. If gene therapy can stop PD disease progression and not just treat symptoms, that is a huge potential benefit for patients. On the other hand, researchers should be extremely cautious when using CRISPR technology as the changes are irreversible.

■ Deep Brain Stimulation

3.1. Traditional Deep Brain Stimulation:

Deep-brain stimulation is the most commonly performed surgical treatment for PD.[14] It is the implantation of wires called electrodes in specific regions of the brain to deliver electrical pulses, helping to regulate abnormal brain signals often found in PD that cause motor symptoms. Since its approval from the U.S. Food and Drug Administration (FDA) in 2002 for Parkinson's, DBS has helped reduce medication needs and motor symptoms.[15] DBS is only available for those who have had PD for 4 years or more and who benefit from medication

DOI: 10.36838/IJHSR819.6

but still struggle with motor control, such as significant "off times" and dyskinesia.[14] DBS usually only works on lessening motor symptoms and involuntary movements such as stiffness, slowness, and tremors, but cannot significantly help imbalance, gait, or any non-motor symptoms. DBS has also been shown to worsen memory, so DBS is not recommended for those with dementia.[14]

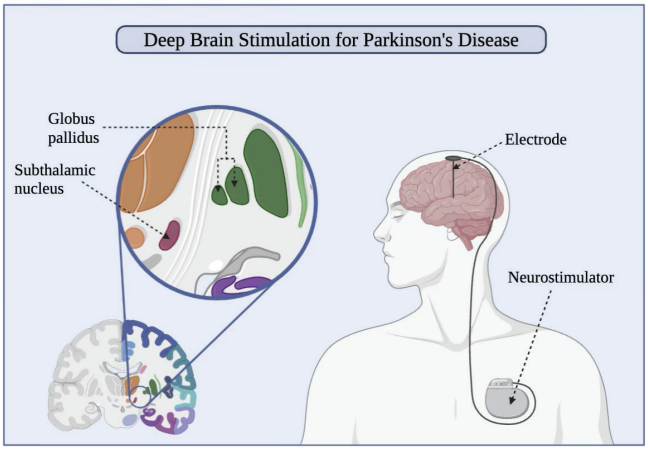


Figure 2: An electrode placed within the basal ganglia is connected to a neurostimulator implanted under the collarbone. Electrical impulses are delivered to the globus pallidus internus and subthalamic nucleus to regulate abnormal neural activity and restore motor control in patients with Parkinson's disease. (Created with BioRender)

In DBS surgery, a surgeon inserts thin electrodes into areas of the brain that control movement, commonly the subthalamic nucleus or globus pallidus interna.[14] While some clinics allow the use of brain imaging to position the electrodes correctly while a patient is asleep, most patients are forced to stay awake and answer questions and perform common tasks. Once the electrodes are in place, it is connected to a neurostimulator implanted under the collarbone, which delivers continuous electrical pulses. The settings of the neurostimulator will later be adjusted post-surgery to optimize symptom control and adjust to new medications. After DBS, most people have been able to decrease the use of medication for PD. However, the downside to traditional DBS is that the right combination of drugs and personalized DBS settings to offer the most benefits and least side effects can take several months to a year to be determined, making it less ideal for PD.[14] Another downside to traditional DBS is that it provides constant electrical stimulation, meaning for those whose symptom severity fluctuates throughout the day, DBS can sometimes be too much or too little, limiting its overall effectiveness.[15]

3.2. Adaptive Deep Brain Stimulation:

One ongoing small study from Nature Medicine tested adaptive DBS, which has the potential to personalize PD symptom treatment.[15] By using custom computer algorithms and artificial intelligence, adaptive DBS can detect symptom fluctuations by monitoring the brain activity of each patient. When it detects a change in symptom severity, it will counteract by delivering a precisely calibrated electrical pulse. Not only can adaptive DBS adjust to symptom severity, but specific motor issues, such as more stimulation for bradykinesia and less for dyskinesia. Moreover, it adjusts to the effectiveness of medication, providing more stimulation when medication wears off and less when it's active. This study enrolled PD patients undergoing deep-brain stimulation in the subthalamic nucleus (STN) and ten PD patients with a globus pallidus (GPi) target, and performed traditional DBS surgery on them.[16] Collecting each patient's brain activity, they used the data to personalize algorithms for adaptive DBS and surgically switched to adaptive DBS.[15] To compare the traditional form of DBS with adaptive DBS, patients switched between those two forms every two to seven days over two months, a total of 1 month spent on each. To keep the study constant, the patients and most of the researchers didn't know which type of stimulation was being delivered, called the double-blinded study, thus eliminating the placebo effect. Every day, patients report their symptoms. This study was able to provide an advanced version of DBS to improve motor symptoms in PD patients and is expected to be completed by 2028. With more clinical testing and widespread technology availability, adaptive DBS could become the new standard for PD motor symptom treatment.15 Because the selection of traditional DBS parameters is often based on trial-and-error experimentation, which could take months, adaptive DBS has given hope to the PD community.

3.3. Connectomic Deep Brain Stimulation:

The next solution is connectomic DBS, a type of DBS to improve tremor control while avoiding side effects by using connectomic imaging models to optimize stimulation of the cerebellothalamic and pallidothalamic tracts.[17] This form uses connectomic imaging models to make more precise parameters for certain disease phenotypes, increasing personalization for each PD patient. In a recent 2025 study sponsored by Duke University, people 18 years or older diagnosed with PD who have had bilateral subthalamic nucleus DBS for at least 3 months and suffer from tremors or motor control are currently being recruited. Similar to the adaptive DBS study, this connectomic DBS study is randomized, double-blinded, and crossover. By comparing pathway-targeted stimulation strategies, this trial has the potential to increase DBS precision, target symptoms without causing side effects, and advance personalized care in Parkinson's disease.[17]

3.4. Deep Brain Stimulation, Conclusion:

Looking ahead, scientists are focusing on expanding clinical trials and refining algorithms. In the future, they hope to make these advanced forms of DBS more widely available. If successful, they could become the standard of care, offering people with Parkinson's more precise, effective, and individualized treatment.

■ Cell Therapy

4.1. Types of Stem Cells Used:

The last therapy discussed in this research paper is cell therapies. Research into stem cell therapies has expanded, showing itself to be a promising solution for PD. Cell therapy is a stem cell-derived treatment aimed at replacing damaged neurons

and restoring dopaminergic function, thus decreasing the severity of motor symptoms.

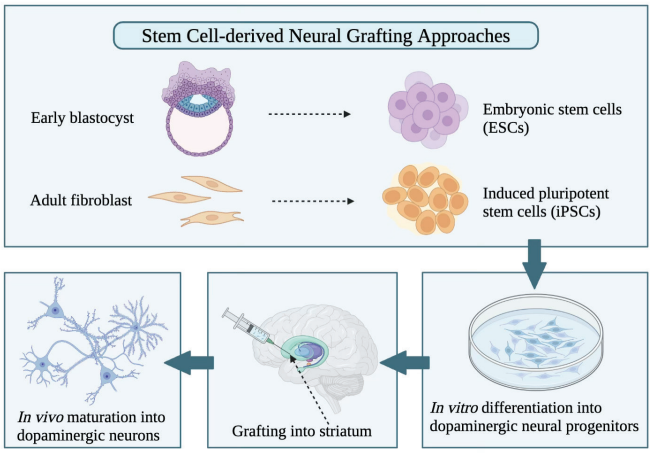


Figure 3: Two approaches for stem cell-derived neural grafting: embryonic stem cells (ESCs) derived from blastocysts and induced pluripotent stem cells (iPSCs) derived from adult fibroblasts. Each cell type is differentiated in vitro into dopaminergic neural progenitors and subsequently grafted into the patient's striatum, where the cells mature in vivo into functional dopaminergic neurons, restoring the patient's lost dopamine production. (Created with BioRender)

Stem cells are undifferentiated cells that are precursors to a specialized cell determined by growth factors.18 One type of stem cell used for PD is embryonic stem cells (ESCs). ESCs are isolated from the inner cell mass of an early blastocyst and generated from excess embryos from in vitro fertilization. Scientists are now able to turn embryonic stem cells into dopaminergic neuron progenitors that closely resemble the authentic nigral neurons lost in Parkinson's disease. However, even with improved models, results are inconsistent. A large study of over 500 grafts showed that success depended on having many tyrosine hydroxylase (TH) positive cells, a key enzyme in dopamine production, in the grafts that contained neural progenitors with caudal midbrain markers such as CNPY1 and EN1. Now, researchers can generate dopaminergic progenitors with much higher purity and reliability. Progress with ESC therapy has been iterative and has led to many new findings in neurodevelopmental biology. Now that scientists have successfully developed dopaminergic neuron progenitors in the lab, the next focus of ESC therapy will be testing clinical efficacy, with many new clinical trials beginning to test this.

Induced pluripotent stem cells are another type of cell often used for PD. iPSCs are created by "reprogramming" an adult somatic cell into a stem cell through the use of transcription factors that induce pluripotency when expressed. Similar to ESCs, the iPSCs can be differentiated into dopaminergic neurons once "reprogrammed." One potential advantage to iPSC-derived grafts is that autologous grafts can be made using a patient's own fibroblast, meaning the neural grafting product would not be suppressed by the immune system, unlike ESC-derived grafts.[18]

4.2. Stem Cells Clinical Trials:

In one 2018 phase I/II clinical trial conducted by Kyoto University Hospital in Japan, six patients aged 50–69 received bilateral transplantation of dopaminergic progenitors from induced PS cells in the putamen.[19] In another recent phase I/II clinical trial published by *Nature* conducted in the U.S. and Canada, twelve participants received transplantation of a human embryonic stem cell (ESC).[20]

Both studies divided the participants in half: one group with a higher dose of cells, the other with a lower dose.[21] All participants were given immunosuppressive medication to prevent their immune systems from attacking the new dopamine-producing cells. Results showed no serious adverse events after transplantation in either study. MRI scans have also shown no signs of tumor formation.

Among the six participants of the iPSC study, four showed improvement on a standard scale used to measure Parkinson's motor symptoms when they were off their traditional PD medication, and five showed improvement when on medication. Additionally, brain scans using a specialized tracer that detects dopamine production showed an average of 44.7% increase in dopamine activity in the putamen, with an even greater increase in those who were on a higher dose of transplanted cells.

The results of the ESC study were similar to the results of the study involving iPSC, as signs of improvement were evident. Brain scans taken 18 months after the transplant showed increased activity in the putamen, indicating that the transplanted cells had survived and were potentially functioning.[21] Moreover, those who received a higher dose of cells showed an average of 23 points higher on the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) when they were off their medication.[20]

4.3. Cell Therapy, Conclusion:

Although these studies show promising results, there are some disadvantages to stem cell therapies.[18] Specifically for ESC therapy, there is an ethical concern regarding cell grafting techniques as they involve the destruction of an embryo, which continues to be a concern for many. Contrastingly, because iPSCs use an adult cell, this makes iPSC therapy also disadvantaged, as adult cells are less differentiated than embryonic cells. Because of multifaceted approaches, it cannot yet be determined which therapy is better without more advanced clinical trials. Another concern regarding stem cell therapies involves the safety of the treatment, specifically the potential for tumor formation, which can form due to graft overgrowth or mutations in the cell culture. However, some recent studies, such as the ones discussed above, have changed their differentiation methods, leading to more refined cell products of increased purity, thus not causing tumors.[18] And while more research is needed to confirm the safety and effectiveness of this treatment, particularly in larger and longer studies, it suggests the possibility that stem cell therapy can provide a solution to PD motor symptoms without causing severe adverse side effects.[21]

DOI: 10.36838/IJHSR819.6

Conclusion

Even though Parkinson's Disease is the fastest-growing neurological condition in the world, because there has been no solution to cure or prevent PD, it costs families and the U.S. government $52+ billion annually. Meeting this challenge necessitates federal research funding for Parkinson's. A $600 million annual federal investment in Parkinson's-specific research at the National Institutes of Health by 2028 could potentially turn mere discoveries into cures, helping to save or improve millions of lives.

Gene therapy offers a promising symptomatic solution for PD for those with a variety of genetic mutations by slowing the progression of symptoms and increasing dopamine production, meaning more motor control. Deep brain stimulation, while still early in clinical development, also holds the potential to be the standard of care for PD patients by offering more precise, effective, and individualized treatment. Lastly, if scientists address safety and ethical concerns, cell therapy can also help PD patients by providing a solution to PD motor symptoms without causing severe adverse side effects.

An estimated 1.1 million people in the U.S live with PD, demonstrating an urgent need for treatments that slow or stop the progression of PD with fewer side effects.22 DBS has the highest potential of being the most accurate PD treatment as it is already FDA-approved for PD and is personalized to each patient, ensuring that each patient receives the care they need. Gene therapy is also promising; however, understanding side effects is critical, as this process is irreversible. Cell therapy has the most barriers, as ESC therapy has ethical concerns over the destruction of an embryo, whereas adult cells in iPSC therapy are less differentiated than embryonic cells, making the therapy less effective. By being extremely cautious with gene therapy, personalizing deep brain stimulation, and finding solutions to the ethical concerns of cell therapy, the treatment of PD is promising. In this review, we discussed the treatments of PD: gene therapy, deep brain stimulation, and cell therapy, and the pros and cons of these treatments. Moving forward, researchers and clinicians should respectively focus on addressing therapy concerns, such as the irreversibility of gene therapy, and closely follow these emerging therapies in order to offer their patients the best possible treatments to improve their quality of life.

Acknowledgments

I'd like to acknowledge Dallin Lowder, PhD, for mentorship, accountability, and answering my questions throughout the research and drafting process. I'd also like to acknowledge the use of BioRender in the creation of Figures 1, 2, and 3.

References

1. Franchina, P. Parkinson's Advocates Rally in D.C., Urge Congress to Boost Research Funding & Ban Parkinson's-Linked Chemical. *American Parkinson Disease Association* [Online] **2025**. https://www.apdaparkinson.org/article/parkinsons-advocates-rally-in-d-c-urge-congress-to-boost-research-funding-ban-parkinsons-linked-chemical/.
2. Fleming, A. Parkinson's Disease, Animation. *Alila Medical Media*. YouTube, Feb 20, 2018. https://www.youtube.com/watch?v=NAfQoviLFR8.
3. Smith, L. and Schapira, A. H. V. *GBA* Variants and Parkinson's Disease: Mechanisms and Treatments. [Online], April 8, 2022. NIH. https://pmc.ncbi.nlm.nih.gov/articles/PMC9029385/.
4. SNCA gene. *MedlinePlus* [Online]. https://medlineplus.gov/genetics/gene/snca/.
5. Current Findings from PD Generation. *Parkinson's Foundation* [Online]. https://www.parkinson.org/advancing-research/our-research/pdgeneration/genetics-behind-pd.
6. Rui, Q., Ni, H., Li, D., Gao, R., and Chen, G. The Role of LRRK2 in Neurodegeneration of Parkinson's disease. [Online], Nov 16, 2018. NIH. https://pmc.ncbi.nlm.nih.gov/articles/PMC6251048/.
7. Axelsen, T. M. and Woldbye, D. P. D. Gene Therapy for Parkinson's Disease, An Update. [Online], Jun 13, 2018. NIH. https://pmc.ncbi.nlm.nih.gov/articles/PMC6027861/.
8. Szunyogh S.; Carroll E.; Wade-Martins R. Recent developments in gene therapy for Parkinson's disease. *American Society of Gene + Cell Therapy* [Online] 2025, Volume 33, Issue 5. https://www.cell.com/molecular-therapy-family/molecular-therapy/fulltext/S1525-0016%2825%2900204-7?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1525001625002047%3Fshowall%3Dtrue.
9. Gene therapy clinical trial for Parkinson's disease. *Prevail Therapeutics* [Online] 2025. https://www.prevailtherapeutics.com/parkinsons-trial/.
10. Phase 1/2a Clinical Trial of PR001 (LY3884961) in Patients With Parkinson's Disease With at Least One GBA1 Mutation (PROPEL). *NIH Clinical Trials* [Online], Nov 20, 2025. https://www.clinicaltrials.gov/study/NCT04127578.
11. Schüle, B. CRISPRi/dCas9 Downregulation of Alpha-synuclein as a Novel Gene Therapy. *The Michael J. Fox Foundation For Parkinson's Research* [Online] **2019**. https://www.michaeljfox.org/grant/crispridcas9-downregulation-alpha-synuclein-novel-gene-therapy.
12. Xiong, Y., Dawson, V. L., & Dawson, T. M. LRRK2 GTPase dysfunction in the pathogenesis of Parkinson's disease. *Biochemical Society Transactions* **2012**. *40*(5), 1074–1079. https://doi.org/10.1042/BST20120093
13. Wojewska, D. N. and Kortholt, A. LRRK2 Targeting Strategies as Potential Treatment of Parkinson's Disease. [Online], July 26, **2021**. NIH. https://pmc.ncbi.nlm.nih.gov/articles/PMC8392603/.
14. Deep Brain Stimulation. *The Michael J. Fox Foundation For Parkinson's Research* [Online]. https://www.michaeljfox.org/deep-brain-stimulation.
15. New Study Further Personalizes Deep Brain Stimulation. *Parkinson's Foundation* [Online] **2024**. https://www.parkinson.org/blog/science-news/personalized-dbs.
16. Starr, P. The Motor Network in Parkinson's Disease and Dystonia: Mechanisms of Therapy. *NIH Clinical Trials* [Online], April 13, 2025. https://clinicaltrials.gov/study/NCT03582891.
17. Connectomic Guided DBS for Parkinson's Disease. *NIH Clinical Trials* [Online], Nov 13, 2025. https://clinicaltrials.gov/study/NCT06618157.
18. Stoker, T.B. Chapter 9: Stem Cell Treatments for Parkinson's Disease. In *Parkinson's Disease: Pathogenesis and Clinical Aspect [Internet]*. Codon Publications, 2018. http://dx.doi.org/10.15586/codonpublications.parkinsonsdisease.
19. Sawamoto, N., Doi, D., Nakanishi, E., Sawamura, M., Kikuchi, T., Yamakado, H., Taruno, Y., Shima, A., Fushimi, Y., Okada, T., Kikuchi, T., Morizane, A., Hiramatsu, S., Anazawa, T., Shindo, T., Ueno, K., Morita, S., Arakawa, Y., Nakamoto, Y., Miyamoto, S., Takahashi, R. and Takahashi, J. Phase I/II trial of iPS-cell-derived

dopaminergic cells for Parkinson's disease. *Nature* [Online] **2025**. https://www.nature.com/articles/s41586-025-08700-0.
20. Tabar, V., Sarva, H., Lozano, A. M., Fasano, A., Kalia, S. K., Yu, K. K. H., Brennan, C., Ma, Y., Peng, S., Eidelberg, D., Tomishima, M., Irion, S., Stemple, W., Abid, N., Lampron, A., Studer L. and Henchcliffe, C. Phase I trial of hES cell-derived dopaminergic neurons for Parkinson's disease. *Nature* [Online] **2025**. https://www.nature.com/articles/s41586-025-08845-y.
21. Two New Trials Explore Stem-Cell Therapy for Parkinson's. *Parkinson's Foundation Science News*, May 8, 2025. https://www.parkinson.org/blog/science-news/cell-replacement.
22. Parkinson's Foundation. Statistics. https://www.parkinson.org/understanding-parkinsons/statistics.

Author

Iris Lee is a senior at Mountain View High School with a strong interest in healthcare. She plans to continue her education with the goal of being a nurse practitioner. In her free time, she enjoys ice skating, badminton, hanging out with friends, and shopping.

DOI: 10.36838/IJHSR819.6

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

A Comparison of Quantum Cryptography Protocols for Securing Internet of Things Devices

Karlo N. Rendic

Mamaroneck High School, 1000 Boston Post Road, Mamaroneck, NY, 10543, USA; knrendic@gmail.com
Mentor: Eric Sakk

ABSTRACT: Internet of Things devices have become a staple of our society, whether that be possessions that you have in your household, such as NEST Smart Thermostats, or industrial and military products that are more widely applicable. Regardless of the context, these items must be secured promptly, given the ever-evolving field of quantum computing. Current cryptographic methods in IoTs rely on fundamental cryptographic principles, adopting an algorithmic approach to implement key distribution between an IoT device and a trusted medium. Research has suggested a new approach to ensure the confidentiality, integrity, and availability (CIA Triad) of data transmitted between devices: Quantum Key Distribution (QKD). Such protocols include Bennett and Brassard 84 (BB84) and Ekert 91 (E91), both of which use quantum principles to secure the key being transmitted. These two protocols can help secure IoTs. This paper analyzed both fundamental Quantum Protocols and examined their implementation in a realistic Implantable Medical Device (IMD) setting. It is found that, although both protocols have advantages, BB84 is the more optimal protocol due to a lower Quantum Bit Error Rate (QBER) over distances of 10 meters than E91, although E91 has a higher detection rate of eavesdroppers.

KEYWORDS: IoT, Quantum Cryptography, BB84, E91.

Introduction

IoT devices transmit highly sensitive data daily and are implemented throughout our lives; however, existing security methods may become insufficient over the next 5-10 years as quantum computing advances. Many IoT security protocols rely on classical cryptographic techniques, such as RSA and elliptic curve cryptography, being more algorithm-based. Quantum computing, however, is specifically designed to solve hard problems such as these efficiently, posing a threat to this widely utilized system.[3,7] However, due to their overall limited power and constraints regarding usable circuitry, the optimization of implementing QKDs is limited. The reason that quantum techniques have been proposed as a solution for key distribution is that anyone eavesdropping on a communication channel will affect the photon distribution or photon status. This is due to the Heisenberg uncertainty principle, where a measurement cannot be made without noticeably affecting the system. Researchers have found methods of applying quantum cryptography, where they present a security blockchain intrusion detection system (IDS) for industrial IoT networks, demonstrating how this can be optimized into industrial environments.[2,4] The issue with this, however, is that industrial IoTs have significantly fewer constraints regarding a multitude of sources compared to IMDs; this thereby limits the applicability of implementation into practical usages.

The main identified issues regarding the feasibility of these implementations have been broad, including items such as data leakage and other attacks that are often overlooked.[3] To defend against attacks such as these, as aforementioned, implementing QKDs has been considered. Some attacks, however, such as long-term device non-updating, cannot be avoided; on the other hand, eavesdropping can be prevented.[3] Eavesdropping, in particular, is a key aspect of the protection that protocols BB84 and E91 protect against–E91 through photon entanglement and BB84 through photon polarization.[3] This paper will directly look at the feasibility, along with the practicality of implementing the two aforementioned methods in IMD, such as Pacemakers and Injectable Insulin pumps; IoTs that need the utmost security. These two protocols have been selected due to their foundational applications into other QKDs, with BB84 being simpler, easier to accomplish, and E91 being more complex, safer, but more energy-constrained. BB84 relies on the transmission of quantum bits (qubits) encoded into different polarization bases, where, when intercepted allow the detection of eavesdropping by enabling the sender to detect it by using measurement properties of the quantum states of the qubits.[3] E91 utilizes quantum entanglement and Bell Inequality violations, where, when intercepted, the sender can utilize Bell's inequality to see if there is a violation; meaning no interception due to quantum properties, while when there is no violation, it means there are no quantum properties–meaning interception.[3]

Methods

Programming tools:

To implement this simulation, we used a Python-based model that leveraged the Matplotlib and NumPy APIs for mathematical calculations and plotting, respectively. This enables us to determine the total amount required for the 1000-bit transmission while also providing an estimate of the demands on the IMDS when implementing these two protocols under varying atmospheric conditions.

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.7

Cryptographic Protocols Tested:

Two distinct cryptographic protocols were used to perform the tests: E91 and BB84. BB84 is a quantum cryptographic protocol that uses quantum mechanics to establish a shared key between a sender and a receiver, while ensuring that an eavesdropper is detected. Information is encoded in two distinct bases, as seen in Figure 1, rectilinear (Z) or diagonal (X). This property allows a sender to detect when a photon's base is measured, as it measures the photons; it creates a discrepancy, allowing any eavesdropper to be detected, creating a quantum bit error (QBE).3 A large number of these errors then lead to an increased QBER, demonstrating eavesdropping.

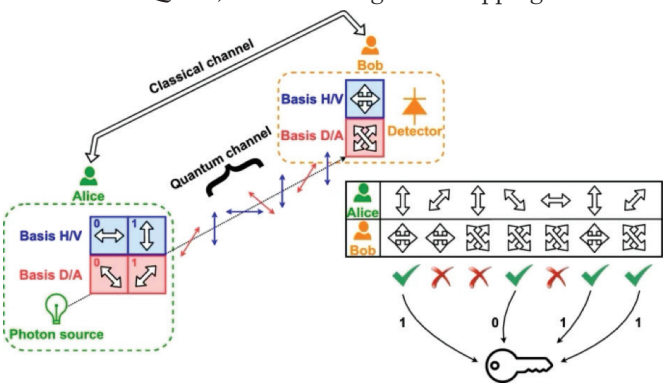


Figure 1: Working Diagram for BB84 Protocol, used along with text for interpretation.[8]

E91, although based on the fundamentals of quantum mechanics, is similar to that of BB8, but it does work using different protocols. As seen in Figure 2, E91 uses Bell's inequality: by sending entangled photons, it can detect eavesdropping. When an entangled photon is looked at, you can utilize Bell's Inequality Theorem to see if it violates the principles of it, then if it does that means that the Entangled Photon still has quantum properties meaning it is not looked at; however, in the opposite case where it doesn't violate it, eaves dropping is present, leading to a more accurate reading than that of BB84.[3] (See Figure 2)

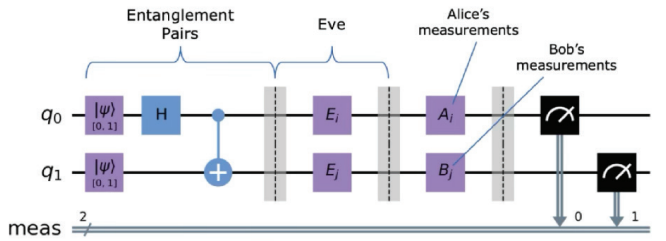


Figure 2: Working Diagram for E91 Protocol, used along with text for interpretation.[9]

Device Models:

The devices that were modeled in this study include IMDs, which include IoTs such as pacemakers and injectable insulin pumps, both extremely resource-constrained. These models utilize realistic constraints to closely copy the conditions under which real IMDs operate. Atmospheric impedance was also utilized in the model, as it affects the energy required to transmit the keys over a distance of 10 meters. We used the wavelength of infrared photons (1550 nm) as a baseline for both of the models–BB84 and E91– which therefore accounts for a more realistic approach to the sending of photons, as most already applicable IoTs, such as Televisions (TVs), utilize this in the communication between remotes and themselves. In this simulation, only a fraction of the original qubits truly reach the receiver, as most are lost, requiring a larger amount of energy utilized to send the same number of qubits, which is our goal of 1000 qubits to create a sufficiently complex key.

Model for Atmospheric Attenuation: Probability of a photon being lost over a distance. α = atmospheric attenuation coefficient and d = transmission distance. Used in BB84 and E91. However, in E91, both pairs need to survive, so that is accounted for by

$p(survive) = (1-p_{atm})^2$.

$$p_{atm} = 1 - e^{-\alpha d}$$

Required Amount of Photons Sent by Sender: To make sure that the receiver receives a target number of usable photons. Used in BB84.

$$N_0 = \frac{N_{target}}{(1 - p_{atm})(1 - p_{Eve})}$$

Calculation For Total Energy Used:

Planck's Relation: used to model infrared photons, where h = Planck constant, c = speed of light, λ = wavelength, in which case wavelength is 1550 nm to simulate infrared photons. Used in both simulations. Used in both to calculate the amount of energy required. (Doubled due to pairs of photons in E91).

$$E_{total} = N_0 E_0$$

Matching Basis for BB84 + E91: Only retains the matching basis to form the key.

$$K = \left\{ i \left| B_A^{(i)} = B_B^{(i)} \right. \right\}$$

QBER Threshold: Used in both simulations to test QBER and to check if it passes a threshold.

Experimental Procedure:

Using a Python program with the application programming interface of numpy, matplotlib, and seaborn, we ran two versions of the simulation, both sending a 1000-qubit-long key. In both, these photons are sent to a receiver, with increasing rates of atmospheric attenuation, along with eavesdropper strength. In the BB84 simulation, the sender encodes a random sequence of bits into photon polarization states according to randomly selected bases; the receiver measures each photon and keeps the corresponding bases. Eavesdropping is then introduced through "eavesdropper strength," which then allows an adversary to intercept and measure the photos being sent, which causes QBE, therefore increasing QBER. Then, using the raw key and the received true key, QBER is calculated; if it exceeds the designated threshold of 11%, the key is subsequently discarded.

The E91 protocol builds on the same framework; however, it utilizes entangled photon pairs. Then the sender and receiver measure their entangled photons with the randomly chosen bases, and only keep those that still contain quantum mechanics. Just as in BB84, the QBER is then calculated regarding the original sent key to the key that is truly used, where the

threshold for discarding a key is the same as in BB84 – 11 %. Just as BB84, this program also accounts for fluctuation values of atmospheric attenuation, looking at different atmospheric conditions, calculated using an atmospheric attenuation coefficient (α), ranging from clean air, 0.1, to contaminated air, 0.30.

Results

BB84:

The results of the BB84 simulation demonstrate the capabilities of BB84 regarding atmospheric conditions; they do not include eavesdropping conditions. In Figure 3, for the case of an adversary, the QBER of BB84 is zero when the adversary is absent (assuming no atmospheric attenuation). This then alters to a QBER of 0.26% when an adversary is introduced, due to the hypothesized disturbance of the quantum states when an adversary tries to read the qubits. Figure 4 reinforces this, as it demonstrates a direct relationship between the adversary's interception strength and QBER: increasing the strength subsequently also increases the QBER, with a max QBER of 0.25 when at full strength, just as the previous bar graph. Figure 5 demonstrates a final quantitative idea: as the attenuation coefficient increases, the number of photons that the sender must send to deliver one thousand photons starts to grow exponentially. This occurs because the higher atmospheric loss causes more photons to be lost, or their path is altered due to other particles in the air. The maximum value of photons required is twenty-five thousand to send one thousand qubits at an atmospheric attenuation coefficient of 0.3.

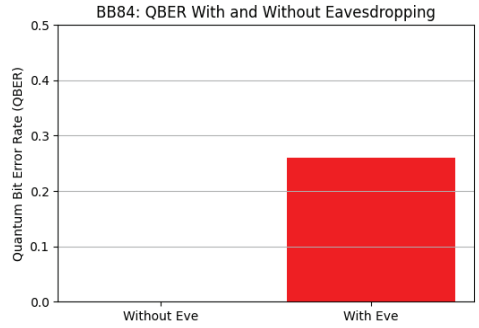


Figure 3: Simulated comparison of Quantum Bit Error Rate with and without adversary (Eve) (1.0 Strength) when using the BB84 cryptographic protocol.

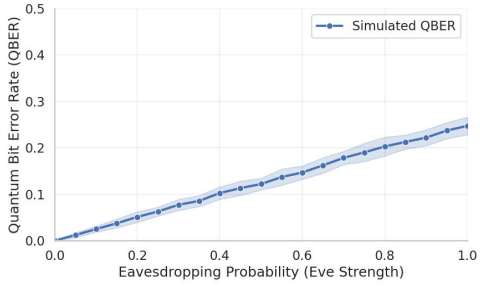


Figure 4: Simulated Comparison of the Eavesdropper Interception Probability v.s. QBER in BB84.

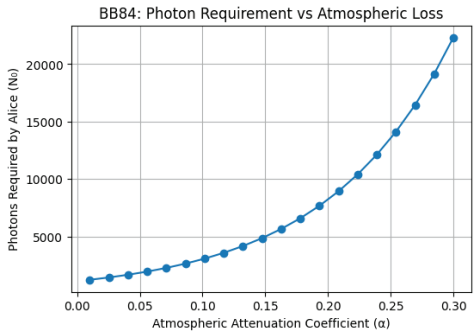


Figure 5: Atmospheric Attenuation Coefficient (α) v.s. Photons Required to Send One Thousand to the Receiver in BB84.

E91:

The results of the E91 simulation, although demonstrating similar patterns to those of the BB84 protocol, show different quantitative values, those of a large caliber. Similar to BB84, when there is no adversary present, the QBER is zero. When this changes to a full-strength adversary, the QBER jumps to 0.5, double that of BB84. In Figure 6, this jump is also seen, where the values of QBER are significantly higher than those of BB84, although both do have a direct relationship between Eavesdropper Strength and QBER, as seen in Figure 7. Another significant change is the difference in photon pairs required to send one thousand over the atmosphere compared to that of BB84, seen in Figure 8. These values are doubled due to the reason that both photons are required to survive the transaction, and if one doesn't, that qubit must be discarded, and that would be registered as a QBE, further increasing the QBER. Otherwise, the graph also increases exponentially as the Atmospheric Attenuation Coefficient increases, just as in BB84.

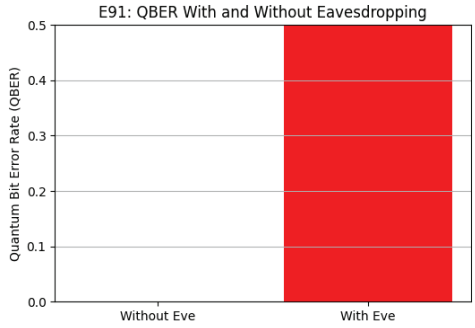


Figure 6: Simulated comparison of Quantum Bit Error Rate with and without adversary (Eve) (1.0 Strength) when using the E91 cryptographic protocol.

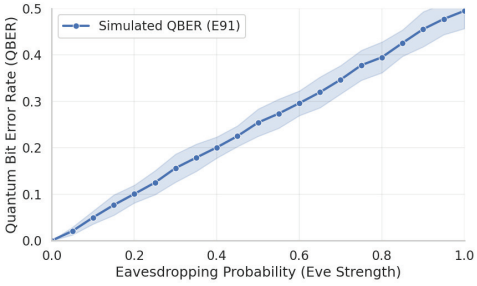


Figure 7: Simulated Comparison of the Eavesdropper Interception Probability v.s. QBER in BB84.

DOI: 10.36838/IJHSR819.7

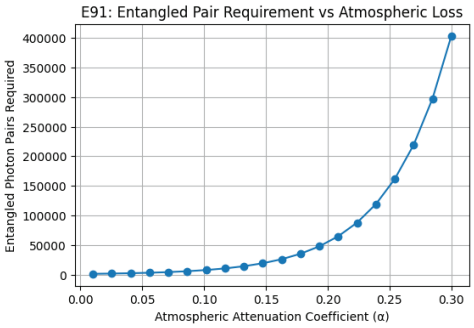


Figure 8: Atmospheric Attenuation Coefficient (α) v.s. Photons Required to Send One Thousand to the Receiver in E91.

Discussion

From our study, it is concluded that although our hypothesis was partially correct, as BB84 is feasible for implementations over IoT networks and E91 is not, it is true that these QKDs could have been more optimized for specific needs, for IMDs. Considering the alignment that is required with these protocols, and with many other technologies that tend to send photons, such as these, which require direct alignment, this is not applicable within certain IoT frameworks. These protocols, specifically that of BB84 due to its smaller photon loss rate, would be more practical in Industrial IoTs and standalone IoTs that don't require movement, such as Pacemakers and Injectable Insulin Pumps.5 Although both protocols are sufficient for detecting eavesdropping regarding the sharing of encryption keys, this is not pertinent due to the required alignment. It is helpful to understand that these protocols are applicable in the frameworks, just not in widely regarded frameworks, such as those mentioned and per se "Smart Home" appliances.

Conclusion

In this paper, we have examined two distinct QKD protocols, BB84 and E91, and, based on our resource-constrained models, concluded that the more stable protocol for IoT applications to be implemented in the near future would be BB84. This is the case because it loses less energy when transmitted through the atmosphere due to atmospheric attenuation, while still maintaining a stable detection rate of eavesdroppers, with a QBER greater than 0.25. This value has been chosen because it is easy to discern whether there has been an eavesdropper, and it is not so low as to be negligible. To reach this conclusion, we have used a Python simulation to ensure that both protocols work as expected and are feasible to implement in a world dominated by IoT devices.

Future Work:

Many items are missing from today's implementations of BB84 and E91. One might look to improve both BB84 and E91 by testing these protocols with other security methods, such as AES, to help create a compromise and get the best of both protocols. This can be further improved by developing a method to send photons in a wider beam, helping extend the range while also increasing the angle of alignment that is required for these protocols, and implementing other post-quantum cryptography methods.[6,7] Other protocols, more complex ones, could also be tested to see their applicability in IoT frameworks as well. Finally, as long as improved engineering comes along in the next 5 - 10 years, using these procedures may become a very real solution to address post-quantum security issues associated with classical key distribution.

Acknowledgments

I would like to express sincere gratitude toward my mentor, Dr.Sakk, and all the time he put in to help me with my research paper. This paper would not have been possible without him.

References

1. Dhar, S.; Khare, A.; Dwivedi, A. D.; Singh, R. Securing IoT Devices: A Novel Approach Using Blockchain and Quantum Cryptography. *Internet Things* **2024**, *25*, 101019. https://doi.org/10.1016/j.iot.2023.101019.
2. Chawla, D.; Kumari, S.; Rathore, R. S.; Mehra, P. S.; Das, A. K.; Kumar, N. Quantum Blockchain for Internet of Things: A Systematic Review, Proposed Solutions, and Challenges. *Comput. Electr. Eng.* **2025**, *126*, 110524. https://doi.org/10.1016/j.compeleceng.2025.110524.
3. Yang, Y.; Lin, Y.; Xiao, J.; *et al.* Feasibility Discussion of Quantum Cryptography for Internet of Things Security: A Literature Review. *Opt. Quantum Electron.* **2025**, *57*, 264. https://doi.org/10.1007/s11082-025-08168-2.
4. Hussain, N.; Li, S.; Hussain, A.; *et al.* Quantum-Aware Secure Blockchain Intrusion Detection System for Industrial IoT Networks. *Sci. Rep.* **2025**. https://doi.org/10.1038/s41598-025-31985-0.
5. Liu, A.; Zhang, Q.; Xu, S.; Feng, H.; Chen, X.-B.; Liu, W. QBIoT: A Quantum Blockchain Framework for IoT with an Improved Proof-of-Authority Consensus Algorithm and a Public-Key Quantum Signature. *Comput. Mater. Contin.* **2024**, *80*, 1727–1751. https://doi.org/10.32604/cmc.2024.051233.
6. Khan, A. A. Securing IoT Communication with the Integration of Quantum Cryptography and Machine Learning. *Int. J. Intell. Syst. Appl. Eng.* **2024**, *12* (3), 4435–4445.
7. Balogh, S.; Gallo, O.; Ploszek, R.; Špaček, P.; Zajac, P. IoT Security Challenges: Cloud and Blockchain, Post-Quantum Cryptography, and Evolutionary Techniques. *Electronics* **2021**, *10* (21), 2647. https://doi.org/10.3390/electronics10212647.
8. Caputo, C.; Simoni, M.; Cirillo, G. A.; *et al.* A Simulator of Optical Coherent-State Evolution in Quantum Key Distribution Systems. *Opt. Quantum Electron.* **2022**, *54*, 689. https://doi.org/10.1007/s11082-022-04041-8.
9. Pedone, I.; Atzeni, A.; Canavese, D.; Lioy, A. Toward a Complete Software Stack to Integrate Quantum Key Distribution in a Cloud Environment. *IEEE Access* **2021**, *9*, 115270–115291. https://doi.org/10.1109/ACCESS.2021.3102313.
10. Pedone, I., Atzeni, A., Canavese, D., & Lioy, A. (2021). Toward a complete software stack to integrate quantum key distribution in a cloud environment. *IEEE Access*, 9, 115270–115291. https://doi.org/10.1109/ACCESS.2021.3102313

Author

Karlo Rendic is an 11th-grade student with a strong interest in problem-solving and analytical thinking. He spends much of his time coding, solving math problems, or working on his next big project.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

What Drives the US Housing Affordability Crisis?

Leonard Valckx

St. Albans School, Mount St. Alban, Washington, DC 20016, USA; lvalckx@yahoo.com

ABSTRACT: Rising house prices and housing affordability are key issues that affect US households and the economy. Previous research already highlighted rising concerns about US housing affordability, and since 2021, housing has become unaffordable for the median US household, as more than 30 percent of their income is spent on housing. This article takes a more in-depth look at US housing affordability at the metropolitan area level and finds a wide variation in affordability across metro areas. Although no single driver fits all, a case study of 7 metro areas finds that elevated interest rates and high building costs play a major role, in addition to strict regulations and some location-specific factors, such as the makeup of the metro area's economy, demographic changes, agglomeration effects, and income inequality. The importance of these factors, especially agglomeration effects, supply costs, and regulatory constraints, is confirmed by a panel econometric study of all 381 metro areas.

KEYWORDS: Economics, Real Estate Markets, Housing Supply and Markets, Housing Affordability, Housing Policy.

Introduction

Rising house prices and worsening housing affordability are key issues that affect US households and the economy. At the national level, the median house price increased by 85 percent, from $222,900 in 2010 to $405,300 by the end of 2025, according to the St. Louis Federal Reserve.[1] Affordability deteriorated, as evidenced by the increase in the cost of homeownership, rising from 30 percent of median income in 2010 to 42 percent in December 2025, according to the Atlanta Federal Reserve.[2] If the annual cost of homeownership exceeds a 30 percent share of the annual median household income, homeownership is considered unaffordable.[3] The latter has been the case since May 2021. Worsening housing affordability adversely affects economic development and growth, as people may not be able to afford to live in an area with higher wages and need to move further out, which may hurt their chances to work there and impede local businesses from hiring workers and growing.[4] Hsieh and Moretti estimated that such spatial misallocation of labor across US cities lowered aggregate US growth by 36 percent from 1964 to 2009.[5]

The housing market and housing affordability are, by themselves, very important for policymaking. Monetary policy decisions by the Federal Reserve may impact housing affordability and housing activity: when the Fed raises interest rates, homeowners with low fixed rates may become more reluctant to sell, thus constraining the housing supply and causing affordability to worsen.[6] In macroeconomic policy terms, more affordable housing translates into greater tax generation, job creation, economic development, and an ability to address inequality.[7] Affordable housing is also a recurrent political campaign issue and a top concern for voters across all age groups—as illustrated in a survey by Redfin on the 2024 presidential election—and by recent policy debates on this issue in the US government and Congress.[8]

Previous research has highlighted rising concerns about housing affordability in the U.S. Anthony noted the chronic shortage of reasonably-priced housing, as about 30 percent of all US households, comprising both homeowners and renters, have been cost-burdened over the past 25 years.[4] Richardson found that in 2012, 78 percent of new and existing home sales were affordable for a typical family based on incomes and prevailing interest rates, but in 2018, only 56 percent of homes on sale were affordable.[9] By 2024, that share decreased to less than 25 percent.[10] Moreover, metro area data reveal wide differences in affordability across the United States, with San Jose, CA, the most cost-burdened market, where 93 percent of a household's income is needed to make mortgage payments, and Decatur, IL, where typical households need to spend only 17 percent of their income for a mortgage in Q2 2025.[11] Glaeser and Gyourko found that the main explanation for this intensifying affordability crunch is that housing cost increases outpace income growth.[12] Some studies also point to the importance of local factors, such as the role of overtourism, out-of-town buyers, and housing supply constraints in mountain towns and coastal areas, and broader macroeconomic conditions, such as elevated mortgage interest rates and economic uncertainty.[13-15]

The main contribution of this article is to examine the evolution of US housing affordability and its main drivers at a granular geographical level. While the majority of past research focused on cross-country comparisons and national or state-level housing affordability, this article analyzes housing affordability at the metropolitan statistical area (MSA) level.[16] This will allow more in-depth investigation of all the local factors underpinning housing affordability and price trends across different geographical parts of the United States, drawing on a comprehensive conceptual framework. This includes the role of government policy, factors that affect the housing supply and housing demand, and household characteristics.

The literature on housing affordability has emphasized several distinct theories, although an all-encompassing theory remains lacking. The major theory of housing affordability focuses on *the supply side*, gauging whether the housing market

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.8

is delivering appropriately priced units. Glaeser and Gyourko found that about 75 percent of US housing markets were priced near or below minimum profitable production costs in 2013.[12] However, subsequent work indicated that the intensity of new housing production dropped substantially over time across the entire US, while housing prices hit historic highs, which suggests that there has been an economically important flattening of the empirical housing supply curve, contributing to unaffordability.[17] *Regulatory theories* argue that construction and zoning restrictions may heighten construction costs and dampen supply.[18] Gyourko, Hartley, and Krimmel found further evidence of large differences in the land use regulatory environment in the US housing market.[19] Baum-Snow and Duranton argue that land use regulation has grown more stringent as there is less land available for development in the most productive and highest-amenity locations.[20] This incentivizes incumbent property owners in the best locations to enact land use regulations to protect their property values, while at the same time prompting a decline in affordability. Such established residents' use of power is referred to in the literature as 'NIMBYism' for the 'not in my backyard' sentiments and has been found to constrain much-needed housing supply.[21] *Growth theories* attribute increased demand to demographic changes and migration. Anthony found that MSAs with high population growth rates from 1990 to 2000 experienced greater increases in the number of housing-cost-burdened households.[22] Gong and Yao showed that demographic changes can account for 40 percent of observed US house price growth from 1970 to 2010, through rising life expectancy, increased urbanization, and international immigration, more than offsetting the negative effect from lower fertility over this period.[23] *Income inequality theories* hold that widening disparities exacerbate housing inaccessibility by causing a concentration of purchasing power among high-income households, which are able to absorb escalating housing expenses and outbid others.[10,24] *Agglomeration theories* contend that larger cities boost worker productivity, increasing housing demand, and *induced migration theories* explore how better amenities attract new residents and contribute to worsening affordability, in the absence of increased supply.[25] Combes and Gobillon reviewed the empirical evidence on urban agglomeration economics and found that average agglomeration effects, measured by the elasticity of productivity or wages with respect to density, are typically in the range of 1–8 percent.[26]

The outline of this article is as follows. First, it presents the data and methodology and some broad trends in house prices and affordability across MSAs. Then, it takes an in-depth look at seven MSAs experiencing different house price and affordability developments to identify the different driving factors, including socio-economic, regulatory, demand, and supply-side factors. The next section provides a quantitative analysis to determine the relevance of the various factors identified from the case studies across all MSAs between 2006 and 2025. The last section concludes and offers suggestions for policy and further research.

Methods

This section describes the data used to assess housing affordability and the main drivers. It also outlines the methodology to assess housing affordability and presents the broad evolution of housing affordability across the 381 US metro areas.

Housing affordability is typically measured by income-based measures, most notably the share of income that households spend on housing.[3] Lenders usually require housing costs to be less than 30 percent of monthly gross income. Anything higher would characterize a household as housing cost-burdened, and housing affordability problems arise if the median household, or a large share of households in a particular region, is housing cost-burdened.[4] An alternative affordability measure is the house price to income (HPI) ratio. Typically, banks consider a good HPI for most homebuyers between two and three, and no more than three to five, but note that this depends on a household's personal situation, including income stability, monthly expenses, expected future earning power, and overall debt-to-income. As noted by Meen and Whitehead, the distinction between cost-to-income and HPI affordability in relation to home ownership reflects an emphasis on, respectively, repayment affordability—i.e., the proportion of income required for mortgage repayments plus other costs— and purchase affordability—i.e., the borrowing capacity and access to credit.[27] Both indicators focus on the economic dimension of affordability and are used widely due to their ease of computation. However, they ignore other aspects of affordability, such as social and environmental criteria, such as evolving quality standards, and neighborhood and transportation issues.[28]

This article compiles house price and housing affordability indicators at the metro area level, using data from Zillow, the Census Bureau, the Bureau of Economic Analysis (BEA), and the Bureau of Labor Statistics (BLS). MSA-level data allow zooming in on specific trends in narrow geographical areas and provide the finest possible breakdown of economic statistics compiled by BEA and BLS. More specifically, Zillow compiles monthly housing affordability data starting in 2012 and house price data from the early 2000s onward for over 900 MSAs, while the Census Bureau's American Community Survey (ACS) publishes detailed annual housing indicators for 508 MSAs from 2006 onward. This survey also collects detailed socio-economic data at the MSA level that can be used to explain cross-sectional differences in MSA housing trends. The Census Bureau also undertakes decadal studies since 1940, which allow for the detection of longer-term trends at the state level, and since 1990, at the MSA level. BEA and BLS compile detailed data on income and household expenditures, employment, GDP, and prices for 384 MSAs starting in the early 2000s. Given the focus on metro area comparisons, this article combines data from these different sources, resulting in a matching data set of 381 MSAs with continuous data from 2006 onwards. Missing data points were backfilled using the variables' development at the state level or the average from other MSAs in the same state. Data for 2020 from the Census Bureau are missing since the census was canceled due to the pandemic, and census data for 2020 are averaged using observations from 2019 and 2021. Table 1 provides more

details on the explanatory variables used to study housing affordability and house prices, drawing on the various theories of affordability, reviewed above, and the conceptual framework from Galster and Lee.[29] The latter identified nine key relationships that shape urban housing affordability and their respective drivers, including not only housing supply, housing demand, household demographics and productivity, but also land, construction and other development costs, labor supply and labor demand, and market expectations.

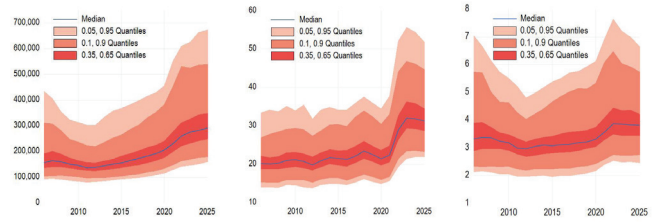
Table 1: Variables included in this study of US housing affordability for each metro area.

Housing affordability	
Zillow cost-income ratio[1]	Share of income that the median household would spend on a newly purchased home that falls in the 35th to 65th percentile range, comprising mortgage payments, homeowner insurance, property taxes, and maintenance costs. Monthly data (since 2012) averaged over the year. The series is backfilled using the Census cost-income trend.
Census cost-income ratio[2]	Selected monthly owner cost (SMOC) divided by median household income, available on an annual basis (2024 data released in October 2025). Census affordability comprises both existing and newly purchased homes, with and without a mortgage, and is not confined to homes that fall within the median range.
House price-income ratio[1,2]	Zillow house price data comprises current-month sales prices, whereas Census Bureau house price data comprises valuations from both current sales and existing home prices based on their tax valuation.
Housing cost-burdened households	Share of households with SMOC exceeding 30 or 35 percent of the median income.
Socio-economic, household, and cost characteristics	
Local housing construction cost index[3,4]	Leland-Saylor national cost index of materials and labor, adapted to local cost by multiplication with the BEA regional price parity for goods (index = 100 in 2010)
Population growth[2,5]	Index for annual population growth (index = 100 in 2010)
Population density[2,3]	Population per square km
Share of non-white households[2]	The complement of the share of the total population that identifies as white, non-Hispanic. Possible indicator to study the effect of race and/or migration
Median age[2]	Median age of the homeowner
Household mobility[2]	The share of households that moved in the last 2 and 3-5 years as a percentage of all households that occupied housing units
Local GDP and sector-specific GDP shares[3]	Real gross domestic product (2017 prices) and shares of specific sectors (government, tourism/accommodation, health, and education)
Productivity[3,5]	Real GDP (in 2017 prices) divided by the labor force.
Commuting time[2]	Average commuting time to work (in minutes)
Gini coefficient[2]	Measure of income inequality (from 0: perfect equality, to 1: perfect inequality)
College population[2]	Share of the over-25-year-old population with a bachelor's or higher degree
Unemployment rate[2]	Unemployment rate for the population 16 years and over
Poverty rate[2]	Percentage of families and people whose income in the past 12 months is below the poverty line
Proxies for housing policy strictness	
Average age of the housing stock[2]	Weighted average age of housing, using total housing units built in each decade starting from the 1930s. A gradual increase over time in the average age of the housing stock may be indicative of metro area housing policies that restrict new housing development.
Vacancy rate[2]	Share of housing units that are not occupied. A low vacancy rate may point to a housing (supply) shortage.
Density of 1-unit structures[2]	Number of one-unit structures (detached and attached) as a share of total housing units. This indicator may capture the (change in) importance of zoning restrictions that favor low-rise, single housing development.

Sources: [1]Zillow Research, [2]Census Bureau's American Community Survey, [3]Bureau of Economic Analysis, [4]Leland-Saylor cost of construction labor/materials, [5]Bureau of Labor Statistics, and author's computations.

To analyze changes in US housing affordability, this article adopts both a qualitative and a quantitative approach. The qualitative approach consists of detailed case studies of seven MSAs to analyze differences in affordability in different parts of the US and examine housing market, socio-economic, and regulatory policy factors. These seven MSAs are selected to cover different geographical parts of the US and to reflect different housing affordability environments. They include coastal areas in the East and West (New York and San Francisco), the capital (Washington, DC), the Midwest/Rust Belt (Detroit), the Pacific Northwest (Bend, OR), and the Sun Belt in the South (Miami, FL, and Austin, TX). These are mostly larger metro areas that account for 15 percent of US households and likely reflect important tendencies that extend to adjacent metro areas. Bend is included to reflect typical affordability challenges of smaller mountain towns and tourism-dependent metro areas stemming from the influx of wealthy second-home buyers.[13,30] The qualitative case study aims to illustrate the relevance of various factors derived from the affordability theories in the previous section. The quantitative analysis in the next section will examine more systematically the relevance of various housing factors for all 381 metro areas over the past 20 years (2006-2025), using a panel regression approach.

Before examining the 7 metro areas in detail, the rest of this section discusses broad trends in house prices and affordability across the 381 metro areas over 2006-2025. House prices increased in phases between 2006 and 2025 and show substantial variation across US metro areas. Between 2006 and 2011, the median house price across US metro areas declined slightly, from $158,000 in 2006 to $137,000, while the compression of house prices in the upper tail of the distribution was substantially stronger, from $434,000 to $$302,000 at the 95th percentile (Figure 1.A). After 2011, the median house price increased strongly, initially gradually until 2019 to $191,000, followed by a steep rise during the pandemic to $261,000 in 2022, and a more modest rise, up to $292,000 in 2025. After 2011, the dispersion in house prices became more pronounced, as evidenced by the difference between the low end and the high end of the distribution, especially after the end of the financial crisis in 2011-2012, and during and after the pandemic, between 2022 and 2025.



A. House prices in US$ B. Housing costs to income C. House price to income

Source: Zillow Research, Census Bureau, author's computations.
Note: Panel figures show fan charts with the median for each variable (blue line) and the distribution around the median in different shades of red, with the middle third (35-65th percentile) in dark red, the 10th and 5th percentile range in medium and light red. Panel A shows the evolution of period-average US house prices for 381 metropolitan statistical areas (MSAs) identified by BEA and BLS using Zillow's smoothed, seasonally adjusted ZHVI series. Panel B shows the Zillow housing cost-to-income ratios for 381 MSAs. Panel C shows the evolution of an alternative affordability metric, house prices to median household incomes.

Figure 1: House price and affordability developments across US metro areas, 2006-2025.

Housing affordability, as measured by the housing cost-to-income ratio, deteriorated most notably in 2022 and 2023 (Figure 1.B). Median housing affordability was well below the 30 percent affordability limit until 2021 and crossed the 30 percent limit in 2023, implying that affordability concerns have affected half of all the US metro areas since 2023. In contrast, 10 percent of all metro areas experienced affordability concerns already starting in 2013, when their cost-to-income ratio passed the 30 percent threshold. Over time, the median housing cost-to-income ratio distribution also shifted upwards, with the range between the lowest and highest 1 percent affordability shifting from 14-33 percent in 2006 to 22-52 percent in 2025, and the middle third of the distribution increasing from 19-22 percent in 2006 to 29-34 percent

DOI: 10.36838/IJHSR819.8

in 2025. This confirms earlier research raising the alarm over housing affordability.[4,31]

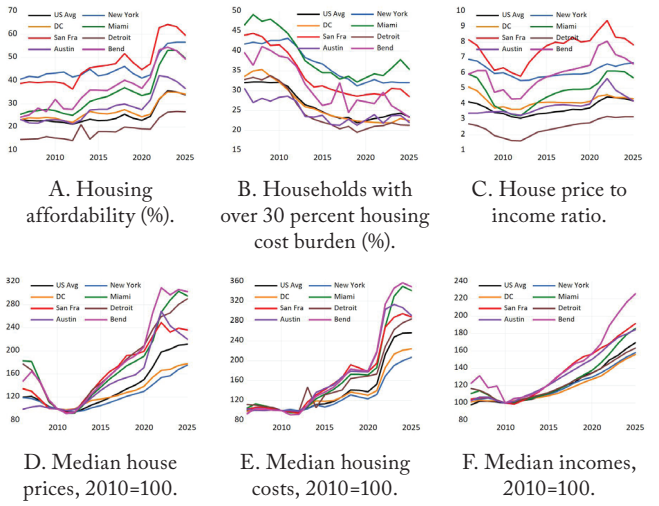
The worsening housing affordability is also reflected in the house price to income (HPI) ratio. The median HPI remained in the 3 to 3.3 range between 2006 and 2020 but sharply increased in 2021 to 3.6 and to 3.9 afterward (Figure 1.C). Thus, the median HPI followed a broadly similar trend as the median cost-income ratio and suggests widespread affordability concerns since 2021 across the majority of US metro areas. Also notable is the sharp decline in the top 5th and top-10th percentile HPI ratios between 2006 and 2012, from 7 to 4.8 and from 5.7 to 4.0, respectively, and their subsequent increase, to 6.1 and 5.3 in 2020, and to 7.7 and 6.5 in 2022, before receding a bit to 6.7 and 5.7 in 2025. Thus, for 10 percent of the metro areas, affordability concerns based on the HPI seem to be even more prevalent and for a longer period of time compared to an assessment based on cost-income. These shifts in the distribution of the HPI and down- and uptrends appear to broadly mirror the acceleration and deceleration of house price increases across these various subperiods, suggesting that incomes developed more slowly. This finding will become more evident in the next section.

Results and Discussion

Housing affordability in seven US metro areas:

This section examines housing affordability in seven MSAs from across the US and local authorities' responses. It finds that affordability issues are due to a combination of factors unique to each metro area.

Housing affordability deteriorated across all seven metro areas, especially after 2020, and notwithstanding some differences, each one has declared a housing affordability crisis. The ratio of housing expenditures to median household income was initially below 30 percent in all metro areas except San Francisco and New York (Figure 2.A). Over time, in all MSAs except Detroit, this affordability ratio crossed the 30 percent threshold, first in Bend in 2010, then Miami in 2012, and during or after the pandemic in the other three MSAs. Housing costs to income in all these MSAs except Washington, DC, and Detroit far exceed the US average. That said, the share of households with a housing cost burden exceeding 30 percent of their income fell between 2006 and 2019 in all MSAs, before increasing again in subsequent years, and appears to be correlated with affordability: the three most unaffordable MSAs (San Francisco, New York and Miami) also have the highest share of households with excessive housing cost burden (Figure 2.B). The alternative affordability indicator, the house price to income ratio, shows a very similar evolution as the housing cost to income ratio, ranking the seven MSAs broadly similarly, and with the same large variation between MSAs, ranging from 2.7 to 8.2 in 2006 and from 3.2 to 7.8 in 2025 (Figure 2.C). This means that the average home in San Francisco costs nearly 8 times the median household's annual income in 2025, whereas it would cost only about 3.4 times a median household's income in Detroit. Despite these differences, authorities in each metro area have declared a housing affordability crisis and announced plans to address this problem (see further below).



A. Housing affordability (%). B. Households with over 30 percent housing cost burden (%). C. House price to income ratio.

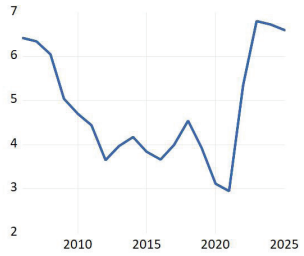
D. Median house prices, 2010=100. E. Median housing costs, 2010=100. F. Median incomes, 2010=100.

Source: Zillow Research, Census Bureau, authors' computations. See also Table 1 for data description. Note: Figures 2.A to 2.C show worsening housing affordability ratios and figures 2.D to 2.F show the underlying elements (indexed with 2010=100), with increases in house prices and cost of housing exceeding households' income growth.

Figure 2: Housing affordability in 7 metro areas.

Across all seven MSAs, the main reason for the deterioration of housing affordability appears to be the inability of incomes to keep up with house prices and costs. House prices and costs on average rose much more than incomes in each of the 7 metro areas, which can be seen from Figures 2.D to 2.F, which index house prices, costs, and incomes with 2010 as a base year. By 2025, house prices rose to around 300 (tripling) in Bend, Miami, and Detroit, and increased the least, by 60 percent, in DC and NY (Figure 2.D). Housing costs followed that trend, or increased even more, by up to 3.5 times in Bend and Miami, and more than doubled in DC and New York (Figure 2.E). Household incomes, on the other hand, increased much less across all metro areas (Figure 2.F). They rose the most in Bend, doubling, and by about 80 percent in San Francisco, Miami, and Austin, while rising by the same amount as house prices in DC and New York, but less than the housing cost increase.

Financial factors appear to play a large role in the housing cost increase. In general, the increase in the cost of housing from 2021 onwards seems related to sharply rising mortgage rates (Figure 3), as interest-sensitive principal and interest payments represent the bulk of monthly housing costs, between 70 and 85 percent in these MSAs. To illustrate further, a 50-basis-point decrease in the 30-year mortgage interest rate, from 6.6 percent (in 2025) to 6.1 percent, would improve affordability by between 1 and 2.5 percentage points across these seven MSAs, and the number of households with unaffordable housing cost burden would fall by about 185,000, consistent with calculations by Zhao.[10] This also ties in with a recent US government proposal to lengthen the maturity of mortgage loans from 30 to 50 years, which would have qualitatively similar effects on affordability, reducing monthly payments, but at the same time lengthening the payback period by another 20 years.



Source: St. Louis Federal Reserve.

Figure 3: The 30-year mortgage interest rate (%).

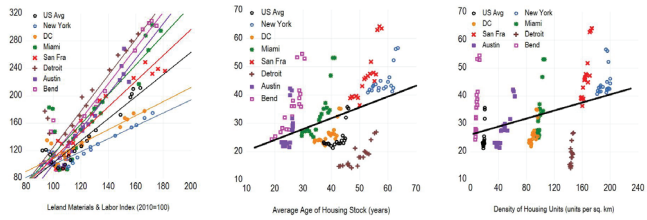
Table 2: Affordability impact of lower interest rates.

				Impact the interest rate drop (-50 bps) on	
	House Price ($)	Cost-burdened households	Cost-Income ratio	Cost-income	HH cost burdened
New York, NY	687,917	2,400,137	56.7	54.6	-88,543
Washington, DC	567,064	551,764	34.3	33.0	-21,932
Miami, FL	476,792	856,924	49.8	48.0	-30,745
San Francisco, CA	1,102,835	506,832	59.7	57.2	-20,840
Detroit, MI	250,514	379,657	26.7	25.7	-14,174
Austin, TX	431,284	236,602	36.7	35.4	-8,490
Bend, OR	661,252	25,157	49.4	47.4	-1,050
Total 7 MSAs	**567,064**	**4,957,072**	**49.4**	**47.4**	**-185,774**
All 381 US MSAs	**356,484**	**27,129,798**	**33.9**	**32.6**	**-1,041,249**

Source: Zillow Research, Census Bureau, author's computations.
Note: Table computes the impact of a 50 basis point (bps) interest rate decline from 2025 levels on the housing cost-income ratio and the associated reduction in housing cost-burdened families. HH = households.

In most of these metro areas, house prices have become high relative to production costs. This confirms the results by Glaeser and Gyourko that these housing markets are generally not delivering appropriately priced housing units.[17] This can be seen from the correlation between our composite measure for the cost of housing, proxied by the Leland-Saylor labor/materials cost index, and house prices, in each of the MSAs (Figure 4.A). Following Lynch, we assumed that the cost of labor and materials represented 50 percent of the house price in 2006, indexed the Leland-Saylor cost series to that level in each metro area, and let it grow over time by the BEA's regional price parity series for goods, which measures differences in goods price levels across metro areas for a given year.[32] For the US as a whole (black dots and line), the cost of supply increased by 63 percent, but house prices more than doubled between 2010 and 2025. For DC (in yellow) and New York (in blue), on the other hand, housing costs track prices relatively well over this period, each increasing by around 70-80 percent. In all other metro areas, the increase in house prices is much stronger than can be justified by the increase in the cost curve, as evidenced by the steeper slopes in Figure 4A.

Strict housing market regulations may explain why construction costs generally pass through imperfectly to house prices. Glaeser and Gyourko explain the generally loose link between costs and prices by the (in)elasticity of the housing supply: in metro areas where housing supply cannot bring on new supply, due to strict regulation of housing construction, the price of housing can be considerably above the minimum profitable production cost, and the shift in demand for housing affects prices more than quantities.[12] This hypothesis is also supported by survey data on the strictness of land use regulations. Gyourko, Hartley, and Krimmel show that a stricter regulatory regime is associated with higher house values, and the most highly regulated markets in the US are situated along either the northeast coast from Boston down through Washington, DC, or the west coast of the country, from Seattle to Portland, San Francisco, and Los Angeles.[19] A relation between regulatory strictness and affordability can be inferred, indirectly, from indicators of the housing stock, which tell how much new supply becomes available over time. Figure 4B shows that metro areas such as Austin, which has a relatively young housing stock and kept its average age constant, appears to have better affordability for most of the time, compared to San Francisco or New York, which have an old and aging housing stock, which suggests that not enough new supply is added to their existing housing stock or restrictive policies, in line with views expressed by real estate reporters.[33,34] Similarly, in Miami, only 4 percent of the metro's housing units have been built since 2010, and housing regulations, especially to increase affordable housing options, seem restrictive and have only recently become responsive to the Miami affordability crisis.[35,36] DC has seen relatively little change in average age of housing, reflecting several large housing developments that have taken place to expand the supply, including downtown and through office-to-residential conversions, which has gained in popularity across the US.[37-39] Detroit, on the other hand, which also has an old housing stock, does not fit this pattern, as it suffered from a foreclosure crisis between 2005 and 2015 which severely limited new construction.[40]



A. House prices and cost of supply (Indices, 2010=100).

B. Affordability and age of the housing stock (percent and years).

C. Affordability and single housing density (percent and units/km^2)

Source: Author's computations, based on data described in Table 1. The line represents linear regression for the cross-sectional data in each figure. The Leland Materials & Labor Index is a proxy for construction costs, and the average age of the housing stock and density of single-family housing units are taken as regulatory housing proxies. House prices are shown to rise more than the Leland housing cost index, and regulatory stringency appears positively associated with lower housing affordability across metro areas.

Figure 4: Housing affordability, construction costs, and regulatory proxies in select MSAs.

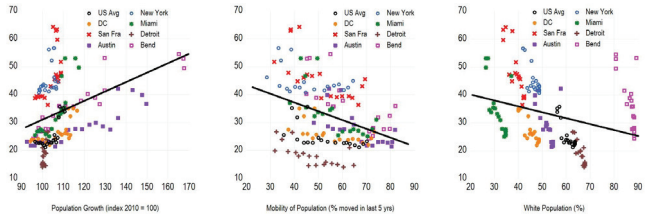
The density of single-family housing units is another proxy for housing regulatory policy, which seems related to affordability. The rationale is that metro areas with high density may be unwilling and/or unable to expand the housing stock in a significant way or may face opposition from existing homeowners when attempting to free up more space (the so-called NIMBY movement), which may put upward pressure on prices and affordability. This can be seen from the contrast in affordability between high-density MSAs such as New York and

DOI: 10.36838/IJHSR819.8

San Francisco versus low-density MSAs such as Austin (Figure 4C). In Austin's case, density bonuses encouraged housing construction, and its affordable-housing program added to the overall supply, whereas metro areas like San Francisco had anti-density policies in place for 50 years.[21,34] In Miami's case, where density bonuses are also available, the housing supply did not expand as local developers felt insufficiently incentivized.[36,41] Bend seems to be an outlier in this relation, as it has a low density but a very high unaffordability, which may be linked to characteristics of mountain towns, which have less habitable space and more supply constraints by the nature of their location, among other factors.

Demographic factors also appear to play a key role in housing affordability in all seven metro areas. Population growth is positively correlated with affordability (Figure 5A). This is consistent with Harvard's Joint Center for Housing Studies 2025 report, noting various demographic headwinds to housing demand, including a slowdown in household formation due to a drop in immigration and less household formation by younger adults, a fall in household mobility, and projected increases in older households due to population aging.[15] Austin and Bend experienced some of the highest population growth and also saw affordability worsen the most, as their median incomes rose less than house prices or costs (Figures 2D-2F). Over time, the share of households that moved between houses has fallen considerably across all metro areas, freezing up the housing inventory-for-sale and thus reducing the supply, which contributes to rising house prices and affordability concerns (Figure 5B). Frey notes that long-distance moves have picked up since 2022, while local moves reached a historic low.[42] The decline in mobility may also partly be attributed to mortgage lock-in, as existing homeowners are unwilling to give up low-rate mortgages at a time when interest rates increased considerably, leading to a spatial lock-in.[6,43] Equally, record-high house prices may keep people from buying their first homes. In turn, when housing is limited, an unwillingness to move out on the part of current residents will hinder outsiders from moving in.[44] There also seems to be some cross-sectional relationship between the racial makeup of these metro areas and housing affordability, as the decline in the share of households identifying as white seems negatively correlated with affordability. This may reflect the large immigration and growth of Hispanic households over the past 20 years, which led to increased demand for housing in each metro area against a constrained supply, leading to higher house prices and worse affordability, in line with Gong and Yao and as projected by Masnick and Di.[23,45] It also follows from the literature on racial disparities in homeownership and housing returns, pointing to worse housing outcomes for non-white households.[46]

Agglomeration effects impact housing affordability across the seven metro areas. Earlier findings on productivity and affordability by Glaeser and Gottlieb and by Van Nieuwerburgh and Weill are confirmed by our data, showing that differences in productivity across metro areas drive differences in affordability (Figure 6A).[25,47] San Francisco and New York stand out for their high GDP per worker. In turn, this allows paying higher wages, which attracts more workers to these places, but also results in more competitive housing demand against an inelastic supply, and thus leads to worse housing affordability. Productivity is also rising in the other metro areas and coincides with decreased housing affordability.

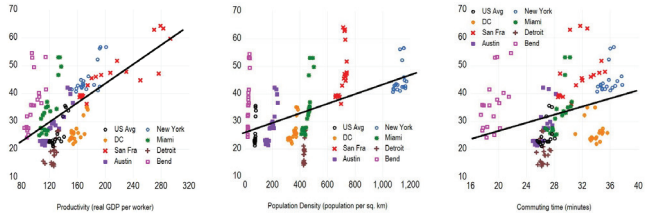


A. Population growth (2010=100) B. Population mobility (%). C. Racial makeup: White (%).

Source: Author's computations, based on data described in Table 1. The line represents linear regression for the cross-sectional data in each figure. Higher population growth is associated with less affordable housing, while mobility and racial makeup (share of white population) appear positively associated with affordability.

Figure 5: Demographics and housing affordability in select MSAs.

Economies of density, measured by population density, are another broad proxy for agglomeration effects, as it influences firms' location decisions.[48] Moving to higher-density areas correlates with worse affordability in our sample of seven metro areas, albeit more on a cross-sectional basis than within each metro area (Figure 6B). Housing affordability problems also seem correlated with longer (willingness to) commute across these metro areas, likely due to the separation of home and workplace, which adds to living costs. Long commutes appear most severe in New York, San Francisco, and DC metro areas, as workers are attracted by higher wages in these metro areas (Figure 6C). However, there has been a reset in commuting in most metro areas after the pandemic, evident from the leftward move in the last three commuting data points (2022-2024), as teleworking took off, whilst affordability concerns continued to rise–possibly due to higher-income households looking for new residences fitting their new lifestyle. The latter may be related to spatial and income inequality, discussed next.



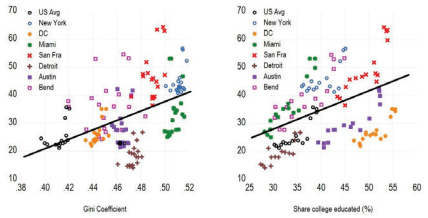
A. Productivity and affordability. B. Population density (per km^2). C. Commuting time (in minutes).

Source: Author's computations, based on data described in Table 1. The line represents linear regression for the cross-sectional data in each figure. Metro areas with higher productivity, higher population density, and longer commuting times appear generally less affordable.

Figure 6: Agglomeration effects and housing affordability in select MSAs.

It is worth noting that these agglomeration effects are reflected in various surveys of livability and city performance. For example, DC is ranked number one on livability in 2025 by Rentcafe.com, based on a combination of socioeconomics, quality of life, and community metrics, far ahead of the next

of our 7 MSAs, New York at 22nd, and San Francisco at 33rd place.[49] A different report, by the Milken Institute, ranked Bend as the fourth-best performer among small cities and Austin as number 9 in the list of large cities, based on metrics from labor market trends, performance of high-tech industries, and shared access to economic opportunities.[50]



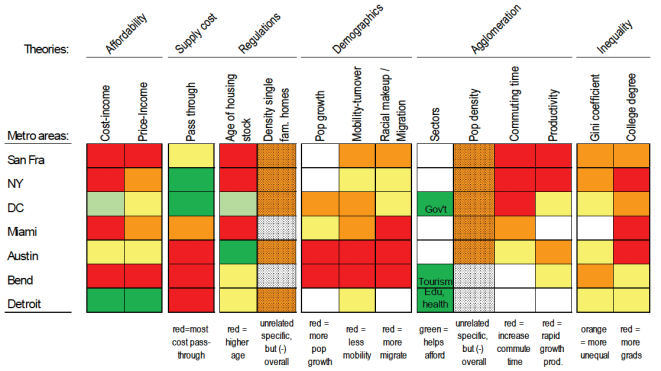
A. Inequality (Gini coefficient). B. College-educated share (%).

Source: Author's computations, based on data described in Table 1. The line represents linear regression for the cross-sectional data in each figure. Higher income and education inequality appear to go hand in hand with less affordable housing across metro areas.

Figure 7: Inequality and housing affordability in 7 MSAs.

Inequality seems intertwined with affordability across all seven metro areas. The cross-section shows a positive relation between Gini coefficients and affordability, although only in Bend, Austin, and San Francisco did income inequality increase noticeably as housing affordability deteriorated. For Bend, this is confirmed by city reports stating that about one-in-six households cannot afford to purchase a home at current market rates, which is attributed to rising income inequality, as the local economy is heavily influenced by service and seasonal jobs that offer low- to moderate-wage jobs that do not keep up with the rising cost of living.[51] In Austin, the rising income disparity reflects an influx of out-of-state households with higher incomes, like tech workers from LA and San Francisco, and the area is disproportionately adding more low- to moderately-paying jobs.[52,53] For San Francisco, this relation is confirmed by Verma, Rinzler, and Zuk, who found that, between 2000 and 2015, increases in house prices and affordability problems were correlated with shifts in where low-income households live.[54] Inequality is also stark in Miami, where housing affordability varies sharply across neighborhoods, with high-income households packed along the region's coastline and in some affluent inland suburbs, and low-income areas sandwiched in between.[35] Another proxy for inequality, the share of college-educated, seems even more strongly correlated with affordability, with marked increases in educational attainment across all seven metro areas, and high house prices crowding out lower-income and lower-skilled households as whole metro areas are auctioned off to the highest bidder.[55]

In sum, a multi-factor explanation seems warranted to explain the housing affordability crisis across these seven metro areas. The above analysis made clear that affordability pressures are different across the various MSAs, and the importance of the various affordability theories/drivers differs across metro areas, as summarized in Figure 8. Some factors have a more severe impact in some MSAs, such as housing cost pass-through in Austin, Bend, and Detroit, demographic factors in Austin, Bend, and Miami, and agglomeration effects in San Francisco and New York. Other factors, such as the single-family housing density regulatory proxy and population density, seem more relevant to explain differences between MSAs than within MSAs, potentially because they are slow-moving indicators, whereas affordability changes have occurred in a shorter time span.



Note: in red = negative effect, green positive, orange and yellow in between, white no effect on affordability. Dashed squares reflect mostly a cross-sectional impact.

Figure 8: Summary of metro area evidence on affordability factors and theories.

In response, each local metro area authority has undertaken efforts to combat its housing affordability crisis (Table 3). A survey of the metro areas' housing authorities reveals that many stress the underproduction of housing units in the past 10-15 years and the fear that unaffordability may drive the workforce away, pointing to negative agglomeration effects. Some financially stronger metro areas have been able to secure funding to invest in affordable homes, while others rely on public-private or philanthropic partnerships. Most metro areas also started to have inclusionary housing mandates, provide incentives for developers to include affordable housing in their projects, or encourage commercial-to-residential conversions.

Table 3: Metro area authorities' plans to make housing more affordable.

Metro area	Main measures to combat unaffordable housing
Austin	Voters approved affordable housing bonds to buy land, build housing, and help with repairs in 2018 and 2023. Austin Strategic Housing Blueprint (2017) set a 10-year goal to create 60,000 affordable housing units through land acquisition and development. Zoning code revisions to enable greater housing density.
Bend	City has a development fee to pay for affordable housing; adopted a system development charge exemption for mixed-income development projects; annexed urban reserve land to build affordable homes.
Detroit	City invests in renovation to improve habitability, has a homeowners' property tax assistance program, adopted a preservation action plan to preserve affordability of housing units, and adopted an Inclusionary Housing Ordinance to develop affordable homes. Philanthropic strategic neighborhood fund finances infrastructure and housing development.
DC	City set a target to create 36,000 new housing units in 6 years (2019-2025), with one-third for low-income earners, including through office-to-residential conversions and development finance for builders. Montgomery County increased financial investments in affordable homes.
Miami	HOMES plan (2022) to preserve/renovate housing, provide mortgage relief, increase the supply of affordable and workforce housing, etc. The Florida Live Local Act offers tax and zoning incentives.
New York	Between 2014 and 2021, NYC financed preservation and new construction of 200,000 affordable homes, passed mandatory inclusionary housing, and various rezoning efforts. New housing plan in 2024 includes significant changes to restrictive zoning rules and a $5 billion commitment to support affordable homeownership, preservation, and infrastructure, while the 2026 housing plan includes the development of public land and 200,000 rent-stabilized units over 10 years.
San Francisco	Developed housing plan (Housing Element) in 2023 for meeting housing needs over the next 8 years (2023-2031), including simplification of permitting and approvals, funding for affordable and large-scale projects, and easing zoning restrictions. Removed barriers for office-to-residential conversions and reformed inclusionary housing and fee system to improve the feasibility of mixed-income projects.

Source: Author's research of metro area housing authority departments.

Quantitative analysis:

This section uses panel regressions to examine housing affordability for all 381 metro areas. This statistical technique finds the common factors driving a specific variable, in this case

DOI: 10.36838/IJHSR819.8

affordability, over time and across a set (a panel) of metro areas. A panel regression can be written in general as:

$$Y_{it} = \alpha + X_{it}\,\beta + \delta_i + \gamma_t + \varepsilon_{it}$$

where Y_{it} is the dependent variable, and X_{it} is a k-vector of regressors (which may include a lagged dependent variable to reflect path dependence), and ε_{it} are the error terms for cross-sectional units observed for $i = 1, 2, \ldots, M$, cross-sectional units observed for dated periods $t = 1, 2, \ldots, T$. The α parameter represents the overall constant in the model, while the δ_i and γ_t represent cross-section and period-specific effects. The β coefficients are assumed to be common across all metro areas and time periods. Least-squares regression is used to estimate these coefficients. To account for the correlation between the residuals and obtain robust estimators, we correct for heteroskedasticity using White two-way clustering of errors.[56] We follow Hendry's general-to-specific estimation approach, starting with an equation that includes all possible regressors, and eliminating one by one those that are insignificant, as judged by t-statistics and other information criteria.[57]

The results show that housing affordability cannot be reduced to a single explanation or single theory. In fact, elements of all five theories of affordability are reflected in the estimated cost-income and HPI panel regression models (Table 4). All variables come with the expected sign, but the LDV in models B and D takes away the explanatory power of some of the variables in models A and C, as expected. The goodness-of-fit, measured by the adjusted R^2, is very high, around 90 percent or more, and the unexplained part of the model is rather small, as measured by the size of the average residual relative to the mean affordability ratio (between 5 and 11 percent). Two variables that measure agglomeration effects are present in all models. Most notably, across all specifications, a higher share of tourism and recreation appears to significantly correlate with unaffordability, likely reflecting the attractiveness of coastal, mountain, and other tourist-rich MSAs for out-of-town buyers.[13,30] In model A, a one-percent increase in the share of tourism in local GDP is associated with a half a percentage point higher housing cost-income ratio. Another agglomeration factor, higher productivity, is also positively correlated with unaffordability in all four models, in line with Baum-Snow and Duranton's model of housing supply.[20] A $10 increase in productivity is associated with a 0.4 percentage point increase in the cost-income ratio in model A.

Greater inequality, measured by the Gini coefficient, is also positively correlated with unaffordability: a 4-percentage-point increase in the Gini coefficient is associated with a half percentage point worsening of affordability. This is in line with Adabre, Chan, and Darko, who associate rising inequality with greater residential segregation arising from big income gaps between high-income earners and middle- and low-income earners, through gentrification or the development of gated communities.[58] Among the demographic factors, higher population growth appears to correlate with less affordable housing across all models, as it leads to increasing demand for housing against a relatively fixed supply, in line with the framework of Galster and Lee.[29] The coefficient size suggests that a 10-point increase in the population index is associated with a 0.4 percentage point increase in the housing cost-income ratio. The cost of supply (Leland-Saylor index) and regulatory proxies (age of housing stock) are statistically significant in three of the four models as well, and their inclusion in the LDV-enhanced models B and D suggests they are equally important to understand housing affordability across US metro areas, in line with Glaeser and Gyourko and Snow-Baum and Duranton.[12,20] A 10-point increase in the supply cost index is associated with the cost-income ratio rising by more than 1 percent, and an increase in the average age of the housing stock by one year would raise the cost-income ratio by about a quarter of a percent.

Table 4: Panel estimation of housing affordability for 381 US metro areas, 2006-2025.

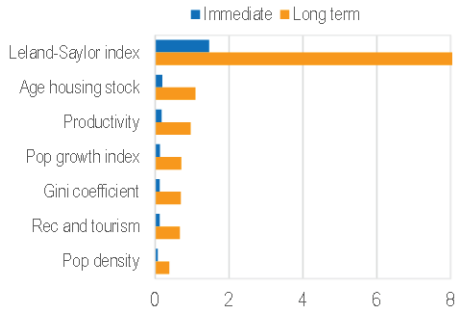
Dependent variable:	Cost-Income ratio		House price-income ratio	
	(A) no LDV	(B) with LDV	(C) no LDV	(D) with LDV
Regressors:	Coef (t-stat)	Coef (t-stat)	Coef (t-stat)	Coef (t-stat)
Leland-Saylor index	0.129 (3.23)	0.063 (2.11)		0.006 (2.59)
Post-2000 housing	0.106 (2.45)			
Age housing stock	0.262 (3.57)	0.060 (2.11)		0.020 (2.92)
Population index	0.042 (1.97)	0.018 (1.91)	0.010 (2.94)	0.003 (2.29)
Mobility	-0.195 (-3.24)			
Edu, health			-0.032 (-2.0)	
Rec and tourism	0.500 (1.94)	0.260 (2.59)	0.052 (1.76)	0.036 (2.41)
Pop density	0.049 (3.33)	0.011 (1.91)		0.003 (2.38)
Commuting time	0.098 (1.87)		0.020 (1.76)	0.008 (1.68)
Productivity	0.042 (2.32)	0.018 (2.17)	0.008 (2.68)	0.003 (2.63)
Gini coefficient	0.126 (2.84)	0.099 (2.17)	0.009 (1.47)	0.013 (3.65)
Lagged dependent variable (LDV)		0.820 (17.1)		0.760 (15.3)
R^2 adj	0.891	0.961	0.911	0.974
N observations	7230	6849	7238	7238
M cross-sections	381	381	381	381
T periods	19	18	19	19
F-statistic	146.3	417.3	183.4	633.3
Average residual	2.6	1.6	0.4	0.2
Mean affordability	24.2	24.3	3.6	3.6

Panel regression estimated with least squares. Standard error and t-statistics are adjusted for clustering.
White two-way cluster standard errors & covariance (degrees of freedom corrected) method is applied.
Cross-section and period fixed effects are included in all models.

One concern with these estimates may be that they are affected by an endogeneity bias, i.e., the fact that no single variable may be strictly exogenous – think of population growth in a metro area being impacted by housing affordability or being jointly determined. This could be solved by using a more sophisticated econometric technique like the generalized method of moments estimator from Arellano-Bond, which estimates the model in first differences and uses lagged (pre-determined) variables as instruments.[59] Although not reported here, such estimation does not qualitatively change the results: housing affordability is still explained by a combination of factors from these different theories, although in more complex ways, with some non-linear effects.

In terms of economic significance, supply cost, agglomeration effects, and regulatory constraints appear most important. While statistical significance establishes whether coefficient estimates for variables are non-zero, economic significance tells which variables have the most important impact on affordability. This article follows Liu and Winegar to calculate economic significance by multiplying the regression coefficients by their within-group standard deviation as the scaling variable.[60] Figure 8 shows that a one standard deviation shock to the supply cost factor raises affordability concerns the most, by 1.5 percentage points immediately and by 8 percentage points in the long run. Agglomeration effects, through the combined effect of shocks to productivity, tourism, and population density, would worsen affordability by 2 percentage points in the long run (by 40 basis points immediately). Regulatory effects, through a shock to

the age of the housing stock, would worsen affordability by 1.1 percentage points in the long run (by 20 basis points immediately). Lastly, a one standard deviation shock to demographics (population growth) and inequality (Gini coefficient) would worsen affordability the least, by 0.7 percentage points in the long run (12-13 basis points immediate impact).



Source: Author's calculation based on Table 4, model B. Impact shows the coefficient x within-group (mean) standard deviation, following Liu and Winegar (2025). Long-term = immediate impact divided by (1-LDV coefficient).

Figure 9: Economic significance of model B variables on housing cost-income ratio (%).

Conclusion and policy recommendations

This article documented the escalating housing affordability crisis across US metro areas over the period 2006-2025, confirming prior academic studies and recent commentaries on this topic, through in-depth qualitative analysis of seven metro areas and an econometric study of all 381 metro areas. The article focused on affordability from a homebuyer's perspective, sidestepping the issue of rental affordability, which deserves a separate analysis. The analysis showed that the magnitude of the affordability crisis varies by metro area, and in some, housing is still relatively affordable, albeit increasingly less so. In both the qualitative and quantitative analyses, the increased unaffordability appears to be caused by house prices and housing costs exceeding income growth, a substantial underproduction of new housing, and a rising housing price-cost ratio. In turn, this may be caused by strict housing regulations, as confirmed by our empirical proxies and by the case study evidence. Against insufficient supply, demand factors such as agglomeration effects, demographic factors such as population growth and race, and income inequality also appear to play a role, as metro areas that offer higher wages, afforded by higher productivity, better amenities, and higher population density grow stronger and create demand pressures between different categories of homebuyers, such as between high and low-income earners, and between residents and out-of-town buyers.

Our research has clear implications for housing policies and for future research. Based on our empirical results, the policy focus should be on expanding supply in a cost-efficient way, such as through cheaper manufactured and modular homes, more commercial-to-residential conversions, and efforts to ease regulatory constraints. Some of the demand-side factors that are found to be important, relating to agglomeration effects, point to the importance of industrial policies to create attractive living conditions for households across all metro areas, so that the incentive to move to so-called superstar cities is mitigated.[55] Finally, macroeconomic and social policies also impact housing affordability through their impact on income inequality and racial disparities in housing. Future research on housing affordability could examine how the various theories fit together in one large theoretical model, e.g., building on Favilukis, Mabille, and Van Nieuwerburgh.[61] Furthermore, changes in housing affordability could be examined over the very long term by exploiting decadal housing surveys going back to 1940, although the data richness may be more limited in the earlier surveys. Finally, an in-depth study of rental affordability, which has worsened alongside homeowners' affordability, and associated policies, would be opportune to accompany the current study, which focused on homeowners' affordability.

At present, housing affordability is receiving wide attention from local authorities, the federal government, and Congress. Many local metro housing authorities have declared a housing affordability crisis and already implemented various measures, including accommodating commercial-to-residential conversions, expanding incentives for developers of affordable housing projects, enhancing mortgage relief, and tapping philanthropic support, etc., as discussed in the case study section. The federal government already supports the US housing market through many initiatives, including housing subsidies, tax relief, mortgage insurance, and housing development incentives, but more may be needed.[62] The US government is now considering declaring a national housing emergency, and its most recent plans from January 2026 include the creation of 50-year mortgages, portable mortgages, directing Fannie Mae and Freddie Mac to buy large amounts of mortgage bonds, allowing people to borrow from their 401(k) to use for a down payment, and banning large financial investors from buying single-family homes.[63] The first four proposals could help lower interest rates and/or increase demand, but in the face of limited housing supply, might cause further house price increases, while the latter may expand the supply of homes-for-sale, but could have negative spillovers onto the rental market. In February 2026, Congress passed some major bills aimed at solving the US housing crisis, including by pushing cities to build more and transferring federal funding away from those that fall behind their target, streamlining federal environmental reviews, relaxing building codes to make it easier to build manufactured homes, and expanding mortgage access.[64] In any case, a solution to the US housing affordability crisis will have to involve a multi-pronged and coordinated response by all levels of government, in coordination with the private sector, banks, and builders, foremost, to tackle supply and regulatory constraints and boost housing demand.[65] The expansion of commercial-to-residential conversion programs seems like a good start to boost supply, possibly in combination with other local efforts and some of the federal proposals listed above, to support both housing supply and demand, in addition to policies that boost household incomes and encourage savings for downpayments.

Acknowledgments

I would like to thank Cesar Martinez-Espinosa from UT Austin/Indigo for guiding me through the research process,

DOI: 10.36838/IJHSR819.8

and the February 2026 UCSB Research Conference participants for their comments. I attest that the ideas, graphics, and writing in this article are entirely my own.

References

1. See St. Louis Federal Reserve Bank for data on the median sales price of houses sold at https://fred.stlouisfed.org/series/MSPUS
2. See Atlanta Federal Reserve Bank for US homeowner affordability data (share of median income) at https://www.atlantafed.org/research-and-data/data/home-ownership-affordability-monitor
3. Joice, P. (2014). Measuring Housing Affordability. *Cityscape: A Journal of Policy Development and Research*. https://www.jstor.org/stable/26326872
4. Anthony, J. (2023). Housing Affordability and Economic Growth. *Housing Policy Debate*, 33(5), 1187–1205. https://doi.org/10.1080/10511482.2022.2065328
5. Hsieh, C.T. and E. Moretti (2019). Housing Constraints and Spatial Misallocation. *American Economic Journal: Macroeconomics* 11 (2): 1–39
6. Van Nieuwerburgh,S. (2026), The Great Revaluation: COVID-19 and the Structural Transformation of the American Housing Market, Chapter prepared for the 2027 International Encyclopedia of Housing and Home (Eds. Acolin, A. and Wachter, S.).
7. Ioannides, Yannis M., and L. Rachel Ngai (2025). Housing and Inequality. Journal of Economic Literature, 63 (3), 916–63. DOI: 10.1257/jel.20251659
8. See Redfin News (2024). Renters Are Nearly Twice as Likely as Homeowners to Say Housing Affordability Impacts Their Pick for President: Survey, at https://www.redfin.com/news/renters-vs-homeowners-housing-affordability-election/
9. Richardson, B. (2019). America's Housing crisis is only getting worse, Forbes, January 31, www.forbes.com/sites/brendarichardson/2019/01/31/americas-housing-affordability-crisis-only-gettingworse/#76649fe9104b
10. Zhao, N. (2026). Pricing in 1.42 million Households: Effects of a 25-Basis-Point Mortgage Rate Cut, National Association of Home Builders, Special Study for Housing Economics, February, 9 pp.
11. Quint, R. (2025). New and Existing Homes Remain Largely Unaffordable in Second Quarter, National Association of Home Builders Economic Research Blog, August, www.eyeonhousing.org/2025/08/new-and-existing-homes-remain-largely-unaffordable-in-second-quarter
12. Glaeser, E.L., and Gyourko, J. (2018). The Economic Implications of Housing Supply. *Journal of Economic Perspectives*, 32 (1), Winter, 3–30, https://doi.org/10.1257/jep.32.1.3
13. On out-of-town buyers and concerns in mountain towns and coastal areas, see Hagedorn, M. (2025). The Housing Crisis in U.S. Mountain Towns. *International Journal of High School Researc*h, 7 (10), 84-98, https://doi.org/10.36838/v7i10.84; Tan, S. R., Wood, J., Yeap, X. W., Li, C., Zhang, C., Wang, Y., Wu, J. (2025). Tourism and housing prices: a meta-analysis. Journal of Sustainable Tourism, 1–29. doi.org/10.1080/09669582.2025.2586540; McNamara, D.E., Smith, M.D., Williams, Z., Gopalakrishnan, S. and Landry, C.E. (2024). Policy and market forces delay real estate price declines on the US coast. Nature Communications 15, 2209, https://doi.org/10.1038/s41467-024-46548-6
14. On interest rate concerns, see Casta, M. (2025). The impact of mortgage rates on the housing market. Journal of Housing Economics 69, September, https://doi.org/10.1016/j.jhe.2025.102079, and Richter, A.W. and X. Zhou (2024). Lower interest rates don't necessarily improve housing affordability. Dallas Fed Economics, April, https://www.dallasfed.org/research/economics/2024/0402
15. On broader macro concerns, see Joint Center for Housing Studies of Harvard University, The State of the Nation's Housing 2025, 48 pp., https://www.jchs.harvard.edu/state-nations-housing-2025
16. Lee, Y., Kemp, P.A. and Reina, V.J. (2022). Drivers of housing (un)affordability in the advanced economies: A review and new evidence. Housing Studies 37(10), 1739-1752, https://doi.org/10.1080/02673037.2022.2123623
17. Glaeser, E.L., and Gyourko, J. (2025). America's Housing Supply Problem: The Closing of the Suburban Frontier? NBER Working Paper No. 33876, May, 44 pp., http://www.nber.org/papers/w33876
18. Quigley, J.M. and Raphael, S. (2004). Is housing unaffordable: why isn't it more affordable? Journal of Economic Perspectives, 18 (1), 191–214. doi:10.1257/089533004773563494
19. Gyourko, J., Hartley, J.S., and Krimmel, J. (2021). The local residential land use regulatory environment across U.S. housing markets: Evidence from a new Wharton index. Journal of Urban Economics, 124, 103337, https://doi.org/10.1016/j.jue.2021.103337
20. Baum-Snow, N. and Duranton, G. (2025). Housing supply and housing affordability, Chapter 6 in Handbook of Regional and Urban Economics, Elsevier, Vol. 6, doi.org/10.1016/bs.hesreg.2025.05.003
21. McNee, G. and Pojani, D. (2022). NIMBYism as a barrier to housing and social mix in San Francisco. Journal of Housing and the Built Environment, 37, 553–573. doi.org/10.1007/s10901-021-09857-6
22. Anthony, J. (2018). Economic prosperity and housing affordability: Lessons from the booming 1990s in the US. Housing Policy Debate, 28(3), 325–341. https://doi.org/10.1080/10511482.2017.1393689
23. Gong, Y, Yao, Y. (2022). Demographic changes and the housing market. *Regional Science and Urban Economics*, 95, 103734, https://doi.org/10.1016/j.regsciurbeco.2021.103734
24. Ge, T. and Kelishomi, A.M. (2025). Income Inequality and Housing Affordability. University of Warwick, Cage Research Centre Working Paper 753, 43 pp.
25. Glaeser, E. L., and Gottlieb, J.D. (2009). The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. Journal of Economic Literature, 47 (4), 983–1028;
26. Combes, P. and Gobillon, L. (2015). The empirics of agglomeration economics, in G. Duranton, V. Henderson and W. Strange, Handbook of Urban and Regional Economics, vol. 15, Elsevier, Amsterdam. http://dx.doi.org/10.1016/B978-0-444-59517-1.00005-2
27. Meen, G. and Whitehead, C. (2020). Understanding Affordability: The Economics of Housing Markets, Bristol University Press
28. Ezennia, I., and Hoskara, S. (2019). Methodological weaknesses in the measurement approaches and concept of housing affordability used in housing research: A qualitative study. Plos One. https://doi.org/10.1371/journal.pone.0221246
29. Galster, G. and Lee, K.O. (2021). Introduction to the special issue of the Global crisis in housing affordability. International Journal of Urban Sciences, 25:sup1, 1-6, DOI: 10.1080/12265934.2020.1847433
30. Favilukis, J. and Van Nieuwerburgh, S. (2021). Out-of-Town Investors and City Welfare. The Journal of Finance, 76, 2577–2638
31. Petach, L. (2022). Income stagnation and housing affordability in the United States. Review of Social Economy, 80 (3), 359–386, https://doi.org/10.1080/00346764.2020.1762914
32. Lynch, E. (2025). Cost of Constructing a Home-2024. National Association of Home Builders, Economics and Housing Policy, January, 11 pp.
33. Shkury, S. (2024). New Housing Policy Could Trigger Development Boom In New York City, Forbes, May 2, online article
34. Brinklow, A. (2025). San Francisco Had An Historic Housing Year, But Not By the Numbers, The Frisc, December 22, https://

thefrisc.com/sf-had-an-historic-housing-year-but-not-by-the-numbers/
35. Florida, R. and Pedigo, S. (2019). Miami's housing affordability crisis, The Miami Urban Future Initiative, Florida International University, 18pp. https://digitalcommons.fiu.edu/record/20361
36. Scotti (2024). Welcome to Miami's housing crisis: How corporations use housing laws to serve shareholders, The [F]Law, November 24. theflaw.org/articles/welcome-to-miamis-housing-crisis
37. Brooks, L., J. Schuetz (2023). Does Housing Growth in Washington, D.C., Reflect Land Use Policy Changes? Cityscape: A Journal of Policy Development and Research, 25 (2), 203-223
38. Eckholm, C., Hadden Loh, T., Meyers, J., and Paynter, S. (2025). Understanding office-to-residential conversion: Lessons from six US case studies. Brookings Metro Research, March 26
39. NPR (2026). From cubicles to kitchens: How empty offices are becoming homes, February 21 www.npr.org/2026/02/21/g-s1-110595/from-cubicles-to-kitchens-how-empty-offices-are-becoming-homes
40. Schuster, S. (2018). Boom and Bust: Detroit's Housing Construction Trends, Data Driven Detroit, July 10, https://datadrivendetroit.org/blog/2018/07/10/boom-and-bust-detroits-housing-contruction-trends
41. Greiner, K.T. (2017). The Dynamics of Housing Affordability in Miami-Dade County: Assessing the Implementation and Impacts of Inclusionary Zoning. South Florida Housing Studies Consortium, April, 69 pp.
42. Frey, W.H. (2023). Americans' local migration reached a historic low in 2022, but long-distance moves picked up. Brookings Metro Update, February 2
43. Hulse, K., Burke, T., Ralston, L., and Stone, W. (2010). The benefits and risks of home ownership for low-moderate income households, AHURI Final Report No. 154, Melbourne: Australian Housing and Urban Research Institute
44. Jones, J.B., Li, Y. and Neelakantan, U. (2025). Why Is Geographic Mobility Declining? Federal Reserve Bank of Richmond Economic Brief, No. 25-19
45. Masnick, G.S. and Di, Z.X. (2003). Projections of U.S. Households by Race/Hispanic Origin, Age, Family Type, and Tenure to 2020: A Sensitivity Analysis, U.S. Department of Housing and Urban Development, Issues Paper on Demographic Trends Important to Housing, 79-123
46. On racial inequality and housing, see Kermani, A. and Wong, F. (2026). Racial Disparities in Housing Returns, American Economic Review, 116 (1), 287-331, DOI: 10.1257/aer.20240327; Reina, V. J., Pritchett, W. E. and Wachter, S. M. (Eds.). (2020) Perspectives on Fair Housing, Philadelphia: University of Pennsylvania Press. https://doi.org/10.9783/9780812297447; and Abedin, S., Benner, L., Khan, J., Rice, L., and Yap, M. (2025). The State of Equitable Homeownership: 2025 Report, The National Fair Housing Alliance, 26 pp.
47. Van Nieuwerburgh, S., Weill, P.O. (2010). Why Has House Price Dispersion Gone Up? The Review of Economic Studies 77 (4), October, 1567–1606, https://doi.org/10.1111/j.1467-937X.2010.00611.x
48. Kopczewska, K., Kubara, M. and Kopyt, M. (2024). Population density as the attractor of business to the place. Scientific Reports 14, 22234. https://doi.org/10.1038/s41598-024-73341-8
49. Dragos, A. (2026). Searching for the best place to move in 2026? These are the most livable metro areas in the US, Rentcafe.com, January 3
50. Switek, M., Smith, B., and Turok, R. (2026). Best-Performing Cities 2026: Resilience in a Cooling Economy, Milken Institute, January, 64pp.
51. Taylor, J. (2025). City of Bend, 2025 State of Housing Report, 60 pp.
52. Austin Community Foundation (2022). Housing affordability in Central Texas: Producing, protecting and preserving, 40 pp.
53. Wadford, J. (2026). On the move: US migration patterns. Bank of America Institute, 22 January, 12pp.
54. Verma, P., Rinzler, D. and Zuk, M. (2021). Rising Housing Costs and Re-Segregation in San Francisco, UC Berkeley's Urban Displacement Project and the California Housing Partnership, 21pp.
55. Gyourko, J., Mayer, C., and Sinai, T. (2013). Superstar cities. American Economic Journal: Economic Policy, 5(4), 167–199. https://www.aeaweb.org/articles?id=10.1257/pol.5.4.167
56. Wooldridge, Jeffrey M. (2002). Econometric Analysis of Cross Section and Panel Data, Cambridge, MA: The MIT Press
57. Hendry, D. F. (2024). A Brief History of General-to-specific Modelling. Oxford Bulletin of Economics and Statistics, 86 (1), 1-20, doi: 10.1111/obes.12578
58. Adabre, M.A., Chan, A.P.C. and Darko, A. (2021). A scientometric analysis of the housing affordability literature. Journal of Housing and the Built Environment 36, 1501–1533, doi.org/10.1007/s10901-021-09825-0
59. Arellano, M. and Bond, S. R. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, The Review of Economic Studies 58 (2), April, 277–297, https://doi.org/10.2307/2297968
60. Liu, Z. and Winegar, A. (2025). Economic magnitudes within reason. Journal of Corporate Finance, 91, April, doi.org/10.1016/j.jcorpfin.2024.102707
61. Favilukis, J., Mabille, P., Van Nieuwerburgh, S. (2023). Affordable Housing and City Welfare. The Review of Economic Studies, 90(1), January, 293–330, https://doi.org/10.1093/restud/rdac024
62. Freemark, Y. and Hermans, A. (2025). How Does the Federal Government Support Housing? An Overview of Federal Supports for Housing Finance, Construction, and Affordability. Urban Institute Research report, October, urban.org/research/publication/how-does-federal-government-support-housing
63. See Parker, W., Picciotto, R., and Friedman, N. (2026), Trump's New Housing Strategy Is All About Boosting Buyers Now, , Wall Street Journal, January 11, www.wsj.com/economy/housing/trumps-new-housing-strategy-is-all-about-boosting-buyers-now-f6a3f6a6 and Tergesen, A., Rubin, R., and Friedman, N. (2026), White House Plan Would Let Americans Tap 401(k)s for Down Payments on Homes, Wall Street Journal, January 16, www.wsj.com/politics/policy/trump-plan-would-let-americans-use-401-k-s-for-down-payments-on-homes-official-says-64eb4375
64. Picciotto, R. (2026). How Congress Plans to Take On the Housing Crisis, Wall Street Journal, February 15, www.wsj.com/economy/housing/how-congress-plans-to-take-on-the-housing-crisis-76bf027c
65. Mehrotra, A., and Ratcliffe, J. (2026). Three Principles of a National Housing Affordability Strategy, Urban Institute, February 26, www.urban.org/urban-wire/three-principles-national-housing-affordability-strategy

Author

Leonard Valckx is a rising senior at St. Albans School in Washington, D.C. His interest in public service led him to examine D.C. housing affordability and housing policies. Through this work, he developed a broader interest in how economic and historical forces influence public policy and hopes to continue exploring these questions in college.

DOI: 10.36838/IJHSR819.8

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Gut Microbiome and Sports Performance: The Symbiotic Relationship

Megan Pan

Mountain View High School, 3535 Truman Ave, Mountain View, CA 94040, USA; megan.m.pan@gmail.com

ABSTRACT: The gut microbiome is a biome of organisms shaped by diet and environment. Humans provide food and shelter to gut microbes, which in turn help with digestion and nutrient absorption. In this review, I discuss the symbiotic relationship between exercise and the increasingly studied gut microbiome. This paper covers: (1) an overview of the gut microbiome, (2) the microbiome's effects on athletic performance, (3) exercise's influence on the microbiome, and (4) future directions and limitations. A diverse microbiome can improve endurance and reduce fatigue, whereas exercise can increase short-chain fatty acid (SCFA) levels and microbial diversity. There is a cooperative relationship between physical activity and the gut microbiome. Increased microbiota diversity benefits sports performance, but misunderstandings persist regarding probiotics. Probiotics are currently viewed as a universal supplement to boost energy and activity. However, this review demonstrates that standardized probiotics are not beneficial because each individual's microbiome is unique. Therefore, this review proposes a transition beyond palliative care to personalized supplementation enabled by omics-based technologies.

KEYWORDS: Medical and Health Sciences, Gut-Muscle Axis, Short-Chain Fatty Acids, Omics-based Precision Nutrition, Microbial Diversity, Probiotics, Athletic Performance.

Introduction

The gut microbiome is a growing subject in research, functioning as a unique microbial ecosystem essential for digestion, immunity, and vitamin synthesis. It is largely influenced by diet, antibiotics, and the environment.[1] The gut microbiota comprises four phyla,[2] which aid in fermentation and metabolite production.[2] "Gut-on-a-chip" technology helps clarify the microbiome's role: it actively guides the immune system toward tolerance.[4] Specifically, these microfluidic models simulate the physical environment of the human intestine, including fluid flow and mechanical stretching that mimics peristalsis (contractions and relaxations of the digestive tract muscle to move food, waste, and water forward). This is relevant because it allows researchers to observe how the intestinal barrier reacts to the physical stress of exercise in a controlled setting, showing how a healthy microbiome maintains tight junctions between cells to prevent systemic inflammation.[5] Gut bacteria prevent inflammatory molecules from crossing the gut barrier and triggering immune responses. By stabilizing this barrier, the gut not only aids digestion but also helps maintain and regulate the immune system.

In this paper, we discuss how the gut microbiome is increasingly recognized as a key factor in athletic performance and health, the symbiotic relationship between exercise and the gut microbiome, and how this relationship can be leveraged to improve athletic performance. The gut microbiome is becoming an increasingly major component of an athlete's capacity. The "gut-muscle axis" suggests that microbial health directly impacts physical output. This axis represents a bidirectional pathway where microbial metabolites influence skeletal muscle function. At the molecular level, this axis functions via the systemic circulation of short-chain fatty acids (SCFAs), which act as signaling molecules that cross the gut-blood barrier to reach muscle tissue.6 This systemic connection is vital for athletes, because impaired gut motility and poor nutrient absorption—which stem from a lack of stable microbiota—can lead to reduced energy. In the context of this review, an "athlete" is defined using the McKinney *et al.* classification, which categorizes individuals based on their training volume (≥10 hours/week) and commitment to formal competition. This encompasses a wide range of athletes, from professionals to recreational athletes, who engage in structured, high-intensity physical activity.[7] Furthermore, the microbiome can protect against physiological stress. High-performance training can cause inflammation, often referred to as a "leaky gut," in which the intestinal barrier is compromised, leading to muscle fatigue and longer recovery times. Instead, a diverse microbiome can strengthen the intestinal barrier, thereby reducing these symptoms.[8] Physical activity increases microbial diversity and SCFA production, regardless of diet.[9] These SCFAs, such as butyrate, act as a critical histone deacetylase inhibitor, which downregulates pro-inflammatory pathways and ensures the intestinal epithelium can withstand the "leaky gut" stress of high-intensity training. This interaction demonstrates that a diverse microbiome is a requirement for peak efficiency. Other bacterial species can recycle waste products into fuel.[10] Therefore, the gut microbiome ensures that nutrients are converted with maximum efficiency.

Despite the numerous benefits of a robust microbiome, probiotics are currently used inconsistently and ineffectively due to the lack of a personalized approach. The goal of this review is to synthesize the current understanding of the gut microbiome and its two-way relationship with sports performance (Figure 1). This paper moves beyond symptom management

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.9

and suggests a proactive, molecularly optimized approach. I will evaluate precision nutrition, in which dietitians shift from generalized probiotic use to individualized, omics-based prescriptions.[11] By identifying specific genetic markers, dietitians can maximize each athlete's metabolic uniqueness.

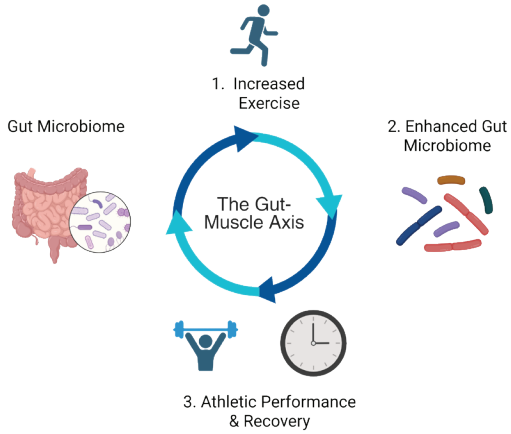


Figure 1: The Gut-Muscle Axis: A Positive Feedback Cycle for Athletes. With increased physical activity, such as running or training, the gut microbial diversity increases. This is indicated by increased production of Short-Chain Fatty Acids (SCFA), including butyrate and propionate. These SCFAs support nutrient absorption, muscle repair, endurance, and reduced inflammation, allowing athletes to continue exercising and further reinforcing the diverse microbial ecosystem. Created using Biorender.

Overview of Gut Microbiome:

The gut microbiome refers to the microbiota (the community of microorganisms) within the gastrointestinal tract (GI).[12] These microbiota aid digestion and nutrient absorption. This leads to fermentation that produces important metabolites, such as SCFAs, lactate, and branched-chain fatty acids. These are essential for athletes.[2] The gut microbiome comprises trillions of microorganisms in the digestive tract.[13] The human gut microbiota contains four phyla: Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria.[2] Specifically, Bacteroidetes and Firmicutes are the two dominant phyla, accounting for over 90% of the gut's bacteria.[14] When examining these phyla within athletic populations, it is important to distinguish between "elite" athletes—those competing at national or international levels— and "recreational" athletes—those who do not meet the 10-hour-per-week training threshold described in the previous McKinney *et al.* definition. This distinction is necessary because exercise volume creates a distinct performance microbiome that may not be present in sedentary or moderately active individuals.[15] In athletic populations, greater diversity of these phyla and a specific increase in the genus Akkermansia have been linked to improved metabolic markers.[4] Many are not host-associated but are distinct microbial species. This microbiome supports various functions, including digestion, nutrient absorption, immune function, regulation of inflammation, and gut-brain communication.[13]

To specifically evaluate the resilience of an athlete's digestive tract under physical strain, researchers created a "human gut-on-a-chip."[16] This miniature device mimicked a human intestine. The chip contained human gut cells and good bacteria, simulating a real digestive system. During this study, they examined gut bacteria, immune cells, and muscle-like movement without harming real-life humans or animals.[16] In the experiment, the transcriptomes of VSL #3, a mixture of 8 probiotic strains (*Lactobacillus casei, Lactobacillus plantarum, Lactobacillus acidophilus, Lactobacillus delbrueckii subspecies bulgaricus, Bifidobacterium longum, Bifidobacterium breve, Bifidobacterium infantis, and Streptococcus thermphilus*), were analyzed to observe their interaction with the host. The probiotics were co-cultured with human gut microbes. After transcriptomic analysis, gene expression patterns in healthy human ileal tissue were identified.[16] The ileum contains more gut microbes than nearby regions and plays a key role in immune function. This shows that the microbiome instructed the immune system to be more tolerant.

Building on this tolerance, commensal bacteria promote the development of anti-inflammatory immune cells, such as regulatory T cells, which then secrete signals, including Interleukin-10 (IL-10), a potent anti-inflammatory cytokine that dampens inflammatory responses.[16] Therefore, due to the microbiome's direct influence, there is a less inflammatory state. Kim *et al.* found that the inflammatory toxin lipopolysaccharide (LPS) failed to breach the gut barrier in the model.[16] LPS is a pro-inflammatory endotoxin found in the cell walls of Gram-negative bacteria that can trigger systemic inflammation if it enters the bloodstream. The microbiome maintains tight junctions between gut cells by producing SCFAs, thereby strengthening the physical barrier.[16] This barrier prevents pro-inflammatory molecules, such as LPS, from entering the bloodstream and triggering the body's immune system. The gut microbiota regulates the balance between pro-inflammatory cytokines, such as Tumor Necrosis Factor alpha and Interleukin-6 (IL-6), and anti-inflammatory cytokines like IL-10 and Transforming Growth Factor-beta, thereby affecting inflammation and the autoimmune system.[17]

The functional impact of this regulation is further clarified by Schrimer *et al.*, who identified a link between the gut and the immune system's inflammatory response: the gut microbiota composition determines the secretion of specific cytokines (chemical signals) into the bloodstream, depending on the bacteria present.[18] During this process, they used *in vitro* simulation with human immune cells (monocytes and lymphocytes) and exposed them to three separate bacterial challenges (*Lipopolysaccharide, Bacteroides fragilis,* and *Staphylococcus aureus*). Then, after the immune cells interacted with the bacteria for a specific period, they measured the concentration of cytokines produced. The results demonstrated that each bacterium had a distinct response.[18] Specifically, certain bacteria induced high levels of Interleukin-1 beta(pro-inflammatory), while others favored an IL-10 response, proving that the host immune profile is highly dependent on individual microbial presence.[18]

The clinical efficacy of these interactions has been demonstrated in trials. For instance, a randomized, double-blind trial showed that daily supplementation with *Lactobacillus helveticus* significantly reduced the duration of Upper Respiratory Tract Infections in athletes by modulating T-cell activity. This effect is likely driven by the strain's ability to enhance Interferon-gamma secretion and stimulate T-cell activation, thereby

buttressing the first line of defense against viral pathogens during post-exercise immunosuppression.[19]

Complementing this defense, clinical trials on healthy, physically active adults using Bifidobacterium animalis subsp. Lactis have shown a reduction in circulating systemic inflammation markers, such as C-reactive protein and IL-6. These findings reinforce the gut microbiome's role in maintaining gut barrier integrity by specifically upregulating tight junction proteins, such as occludin and zonulin. These actions prevent the translocation of endotoxins, which typically trigger systemic inflammation, in highly active individuals undergoing heat or mechanical stress on the gastrointestinal tract.[20]

Gut Microbiome's Role in Athletic Performance:

The gut microbiome plays a key role in sports. A more diverse microbiome often enables higher performance (Figure 2). To illustrate the necessity of these microbes, one study compared germ-free (axenic) animals with conventionally raised animals.[12] The germ-free (GF) animals had less body fat, reduced gut motility, and showed multiple behavioral differences.[12] These axenic animals were completely free of living microorganisms. In Kuziel and Rakoff-Nahoum's experiment, the microbiome caused specific innate and adaptive immune responses.[12] The germ-free mice had a lower basal metabolic rate (BMR), which is neither positive nor negative. A lower BMR can lead to longer lifespans because it requires fewer calories. However, it can also increase the risk of faster weight gain and other health issues. This finding contradicts the initial results, as the experiment found the mice had gained less fat, even though a lower BMR can increase weight.

This study clarifies that when germ-free mice remain lean despite a low BMR, it means that, without a microbiome, an organism is an inefficient nutrient processor. It cannot efficiently use the energy from its diet, which is an immense disadvantage for high performance. While these axenic models primarily highlight fundamental metabolic pathways, the inability to process nutrients efficiently suggests that, without microbial aid, an athlete cannot effectively fuel the high caloric demands of training. The performance gap is further illustrated when comparing GF mice to specific pathogen-free (SPF) mice. SPF Mice possessed a diverse microbiota dominated by *Bacteroidetes* and *Firmicutes* and consistently showed greater endurance on the treadmill.[21] This is due to the microbiota's ability to support skeletal muscle uptake of glucose and maintain antioxidant enzyme activity, which are severely limited in germ-free models.

Gut motility refers to the movement of food and waste through the digestive tract.[22] Impaired motility reduces nutrient absorption. This fundamental link between the microbiome and metabolic efficiency has been demonstrated in human trials. Athletes supplemented with a high-dose blend of *Bifidobacterium* and *Lactobacillus* strains experienced an increased VO_2 max and an extended exercise duration before failure, demonstrating that optimized microbial activity directly translates into measurable aerobic gains.[23] Furthermore, professional cyclists who used a similar mixture of *Bifidobacterium animalis* and *Lactobacillus hevea* reported significantly less nausea and vomiting, suggesting that the microbiome supports performance by maintaining the physical integrity of the GI tract under the stress of training.[24]

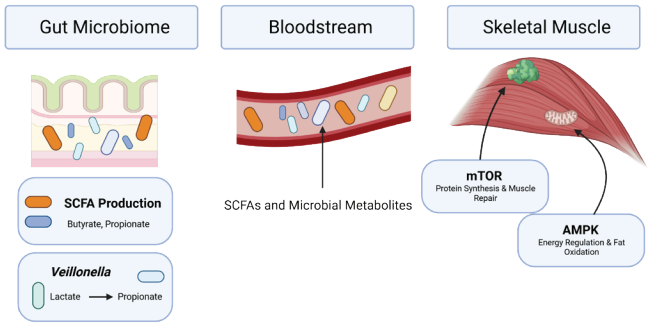


Figure 2: The Gut-Muscle Axis in Athletic Performance. The gut microbiome produces Short-Chain Fatty Acids, such as butyrate and propionate, which cross the intestinal barrier and enter circulation. Additionally, Veillonella, a Gram-negative bacterium found in the gut, metabolizes lactate, produced during exercise, into propionate. These metabolites enter the bloodstream and reach skeletal muscle, where they influence metabolic pathways, including mechanistic target of rapamycin-mediated protein synthesis and muscle repair, and adenosine monophosphate-active protein kinase-mediated energy metabolism and fatty acid oxidation. Created using Biorender.

These findings indicate that athletes with a less diverse gut microbiome may experience poorer athletic performance. If reduced gut motility impairs nutrient absorption, maintaining microbial diversity will be critical for metabolic efficiency and energy availability during high-intensity training. Furthermore, a diverse microbiome strengthens the intestinal barrier, thereby reducing systemic inflammation that often leads to muscle fatigue and prolonged recovery times.[25]

By connecting the nutrient-processing seen in animal models to the increased aerobic capacity in human trials, it is clear that a diverse microbiome optimizes energy conversion. This conversion is primarily driven by SCFAs—small, lipid-like carboxylic acids produced during the fermentation of non-digestible carbohydrates.[26] SCFAs not only provide a supplemental energy source but also act as signaling ligands that enter the bloodstream and bind to receptors in skeletal muscle to regulate fatty acid oxidation and glucose homeostasis.

This relationship between diversity and metabolic output is further supported by a study by Wiley Barton *et al.*, in which athletes had higher fecal metabolite levels—such as SCFAs, propionate, and butyrate—than control (sedentary) groups.[27] Metabolism performs four key functions for cells: generating ATP, converting nutrients, converting simpler structures into macromolecules, and aiding cellular functions.[28] Therefore, with better metabolic regulation, the cell grows only when necessary. Additionally, the enzyme mechanistic target of rapamycin (mTOR) catalyzes anabolic processes by regulating protein synthesis and degradation. In contrast, adenosine monophosphate-activated protein kinase (AMPK) catalyzes catabolic processes by breaking down fats and sugars (Figure 2).[28] The gut-muscle axis influences this balance because SCFAs such as butyrate can inhibit histone deacetylases, effectively upregulating the pathways required for muscle protein synthesis via mTOR. Simultaneously, propionate serves as a precursor for gluconeogenesis, helping maintain energy levels

DOI: 10.36838/IJHSR819.9

and preventing the overactivation of AMPK, which can lead to muscle breakdown during prolonged endurance tasks.[29] When these processes are accelerated, athletes can more easily break down nutrients and build muscle. Barton *et al.* suggest that the high levels of SCFAs found in athletes act as primary mediators between the gut microbiome and cellular metabolic pathways.[27]

Specifically, SCFAs such as butyrate and propionate travel through the bloodstream to skeletal muscle, modulating the balance between mTOR and AMPK. This is exemplified by specific bacteria, such as *Veillonella*, which metabolize exercise-induced lactate into propionate, thereby directly fueling endurance performance.[10] This regulation provides a stable energy source and enables microbial metabolites to more efficiently activate mTOR for muscle repair while activating AMPK to optimize fat oxidation and energy production. As a result, the microbiome links nutrition and exercise, ensuring that ingested energy is converted into physical power. The synthesis of molecular signaling (mTOR/AMPK) and clinical performance outcomes (VO_2 max/fatigue resistance) confirms that the microbiome is a non-negotiable factor in athletic endurance. This indicates that metabolic speed in athletes is determined not solely by training but also by the gut microbiome. Ultimately, these findings shift from a solely external focus on sports performance to an internal focus on microbiome management, suggesting that an athlete's ability is supported by their microbial health.

How Exercise Influences the Gut Microbiome:

While the microbiome supports exercise, exercise has likewise been shown to alter the gut microbiome. Hsu *et al.* compared germ-free (GF) mice with specific pathogen-free (SPF) mice.[30] They had each group of mice perform exercises, including endurance swimming and treadmill running. In these tests, SPF mice exhibited greater exercise capacity and endurance than their GF counterparts.[30] This suggests that a diverse microbiota plays a prominent role in exercise and may be required for optimal performance.

In a study by Mika *et al.*, the authors investigated whether exercise (wheel running) at a juvenile stage (three weeks old) would produce a more diverse microbial community than initiating exercise during adulthood (ten weeks old).[31] Rats were placed in cages with running wheels, providing voluntary wheel access. In their experiment, they found that initiating exercise during the juvenile stage was associated with increased Bacteroidetes and decreased Firmicutes.[31] This Bacteroidetes-to-Firmicutes ratio is often associated with a lean phenotype and improved metabolic health. These microbial groups are crucial because they extract energy from food through fermentation, enabling efficient calorie absorption. The Mika study further reveals that early-life exercise specifically promotes the abundance of *Proboscibacter* and *Lactobacillus* species, creating a microbial signature that enhances the host's metabolic plasticity into adulthood.[32] Consequently, this combination overall indicates a lean phenotype, which is widely associated with regular exercise.

Furthermore, this metabolic shift can be linked with increased SCFA production. SCFAs are signaling molecule that triggers intracellular signaling cascades when bound to specific cells, such as by activating membrane receptors and directly modulating enzymes and transcription factors.[33] This is particularly important because SCFAs such as butyrate provide the primary energy source for colonocytes—specialized epithelial cells that line the lumen of the large intestine—therefore strengthening the gut barrier. This overall indicates that voluntary aerobic exercise positively influences gut health.

In a study by Clarke, S. F. *et al.*, 40 elite male professional rugby players were recruited.[4] They collected fecal and blood samples and extracted DNA to use for the experiment. They found that the alpha-diversity of elite athletes, a measure of the variability of microbial species within a single fecal sample, was significantly higher than that of the high-BMI group and the control groups (sedentary individuals). Alpha diversity was assessed to quantify richness and evenness within the sample. The researchers concluded that higher alpha diversity is associated with better health outcomes.[4]

Adding to the list of benefits, Candace R. Longoria *et al.* established that exercise is among the few lifestyle factors that can independently increase SCFA levels.[9] A nutritious diet can act as fuel, signaling to SCFA-producing bacteria. They found that physical activity can eliminate metabolic waste products, such as trimethylamine N-oxide (TMAO). After ingesting foods high in choline or Lecithin (red meat, eggs, dairy, etc.), certain bacteria in the body ferment them into a gas called Trimethylamine (TMA). This gas is transported to the liver, where an enzyme converts it to TMAO.[9] Exercise reduces the abundance of high-TMA-producing bacteria, thereby reducing the amount of raw TMA gas secreted into the liver. Furthermore, exercise has been shown to downregulate FMO3 activity. This ensures that TMAO production would be slower and TMAO concentration would drop, regardless of diet changes.[9]

Parallel to these findings in elite athletes, there was also an increase in SCFA producers among non-athletes with increasing physical activity.[34] *Faecalibacterium*, a SCFA-producing bacterium, showed higher population levels in relatively active individuals.[35] This is critical because SCFAs can reduce inflammation through signaling pathways, regulate immune function, and modulate obesity.[36]

Future Directions and Limitations:

Probiotics are microorganisms that modulate the microbiome and confer benefits on the host, thereby supporting gut health after antibiotic therapy.[37] Athletes often use them to improve markers of oxidative or inflammatory stress. However, probiotics have not been shown to improve performance.[2] In a group of 24 recreational runners, they were given probiotic supplements (*Lactobacillus acidophilus, Bifidobacterium bifidum*, and *Bifidobacterium animalis subsp lactis*) 28 days before a marathon. Compared with the control group (recreational runners who did not receive probiotic supplementation), the experimental group experienced less severe gastrointestinal symptoms and a modest decrease in average speed. Unfortunately, the finishing times did not differ significantly.[2] This

finding indicates that probiotics did not affect performance but rather helped minimize physical symptoms that would otherwise impair a runner's experience.

Since probiotic effects are highly strain-specific, their efficacy must be evaluated at both the strain and species levels rather than in a broad category.[38] Different strains have different effects on elite athletes. In a study testing *Lactobacillus rhamnosus* (*L. rhamnosus*) strain ATCC 53103 in marathon runners, Kekkonen *et al.* found that the probiotic had no effect on the total incidence or number of respiratory infections, nor did it reduce the number of GI symptom episodes.[39] However, the duration of GI episodes was reduced. These GI symptoms were defined as flatulence, abdominal pain, loose stools, or diarrhea, and participants reported their severity on a scale.[39] Conversely, Shing *et al.* had male runners use a multistrain probiotic supplement (*Lactobacillus acidophilus, L. rhamnosus, Lactobacillus casei, Lactobacillus plantarum, Lactobacillus fermentum, Bifidobacterium lactis, Bifidobacterium breve,* and *Streptococcus thermophilus*).[40] This study concluded that GI discomfort was minimally reduced, but supplementation significantly increased running time to fatigue.[40] These discoveries indicate that different probiotic strains affect athletes differently: *L. rhamnosus* had a negligible effect on marathon runners,[39] whereas multistrain supplementation had a significant effect on male runners.[40] Given that the *L. rhamnosus* strain does not exhibit a significant effect, combining multiple strains might improve athletic performance. This raises a critical question for future research: whether there is an optimal threshold for strain diversity or a point beyond which adding more strains no longer provides additional benefit.

In response to these challenges, precision nutrition, informed by omics technologies and gut microbiome research, offers a new approach to enhancing athletic performance through individualized interventions. Penggalih *et al.* found that precision nutrition can optimize athletic performance by using omics technologies (nutrigenomics, metabolomics, proteomics, and transcriptomics) to dive into an athlete's molecular and metabolic nuances.[11] By identifying specific genetic markers, doctors can create appropriate diets to potentially reduce injury risk or increase endurance. For example, metabolomics can help identify metabolic responses to exercise, aiding dietitians in developing recovery strategies.[11] As shown by Li *et al.*, the future of sports nutrition lies in sports-specific probiotic strategies.[41] This evolution highlights a shift from generalized supplements to a sharper focus on the "gut-muscle axis," where specific microbial metabolites directly influence muscle recovery and energy availability. This research relies heavily on multi-omics, as stated in Penggalih *et al.*,[11] to bring together "responders" and "non-responders." By combining metagenomic screening with metabolomics, future studies can identify which probiotic strains would benefit an athlete. Ultimately, researchers should go beyond symptom management and work toward identifying molecular thresholds at which strain diversity and personalized nutrition plans ensure measurable improvements in athletic power and endurance.

To achieve this, future reserach my implement longitudinal monitoring frameworks that utilize wearable biosensors and ingestible "smart pills" to track real-time microbial fluctuations during high-intensity phases.[42] By mapping these fluctuations, researchers can identify windows of vulnerability in the gut barrier where targeted interventions may be more effective. Furthermore, interventional studies should transition from traditional randomized controlled trials to randomized, blinded, multiple-crossover clinical trial designs that prioritize the individual's unique baseline microbiome and metabolic responses over broad population averages.[43] This will ensure the data is accurate, and any benefits will not be hidden as they might be in a larger group. To incentivize this work, the scientific community should establish open-access "Performance Microbiome" biobanks.[44] These platforms can allow for the sharing of multi-omics data across institutions, encouraging the validation of strain-specific claims through large-scale meta-analyses and reducing the financial burden on individual researchers. Moreover, by encouraging and rewarding researchers for debunking, the field of microbial research will maintain its integrity.

However, knowledge of the gut microbiome is limited due to inconsistent research.[45] Many athletes suffer from gastrointestinal infections due to prolonged, intense exercise.[46] O'Brien believes that this inflammation is due to changes in the microbiome.[46] Although the use of probiotics is increasingly common, the American Gastroenterological Association states that assessing their benefits is difficult due to differences in methodologies, specimens, and participants.[45] For instance, although the *L. rhamnosus* strain had little effect on performance,[39] multiple-strain supplements have been shown to decrease fatigue time.[40] Additionally, Kekkonen *et al.* included participants of both genders, with a sample of 114 marathon runners,[39] whereas Shing *et al.* tested only 10 male runners.[40] This discrepancy underscores that athletes cannot rely on standardized probiotics, as few studies have demonstrated their benefits.

■ Conclusion

This review of the intersection of sports science and microbiology confirms that the human gut is not merely a digestive organ but an increasingly dominant component of an athlete's performance. As shown by the distinct physiological advantages of axenic versus conventionally raised animals, a diverse microbiota is crucial for efficient nutrient absorption and for regulating systemic inflammation.[12] The relationship is symbiotic: a diverse microbiome fuels endurance by producing metabolites such as SCFAs,[10] while physical activity enhances microbial richness and flushes metabolic waste, such as TMAO.[9] This is significant because it highlights how athletes should be more mindful of their gut microbiome through their diet and environment.

Despite this association, the current application of probiotic supplementation remains inconsistent. While single-strain probiotics were ineffective at improving runners' endurance,[39] the multi-strain formula provided fatigue-reducing benefits.[40] These disparities highlight that standardized supplementation is no longer a viable solution for future athletes. Instead, dietitians and researchers should prioritize precision nutrition over

DOI: 10.36838/IJHSR819.9

symptom management. This can take the form of omics-based technologies, such as nutrigenomics and metabolomics, used to optimize athletic performance through molecular individualization.[11]

Ultimately, while exercise has been shown to benefit gut health, further optimizing performance and endurance requires mapping and manipulating the microbiota at the molecular level. With this manipulation, athletes are not limited to generalized estimates; instead, they can use data and molecular studies to identify specific strains and effective doses. There needs to be an interdisciplinary effort among sports nutritionists, immunologists, and data specialists to translate omics into real-world solutions. This approach will ensure that research is inclusive and diverse, including underrepresented groups. Future research should focus on developing personalized probiotic supplementation tailored to the person's microbiome.

Acknowledgments

I would like to acknowledge Dr. Daniel Kim for his support and guidance throughout the writing of this paper.

References

1. Van Hul, M.; Cani, P. D.; Petitfils, C.; De Vos, W. M.; Tilg, H.; El-Omar, E. M. What Defines a Healthy Gut Microbiome? *Gut* **2024**, *73* (11), 1893–1908. https://doi.org/10.1136/gutjnl-2024-333378.
2. Clauss, M.; Gérard, P.; Mosca, A.; Leclerc, M. Interplay Between Exercise and Gut Microbiome in the Context of Human Health and Performance. *Front. Nutr.* **2021**, *8*, 637010. https://doi.org/10.3389/fnut.2021.637010.
3. Sarita, B.; Samadhan, D.; Hassan, M. Z.; Kovaleva, E. G. A Comprehensive Review of Probiotics and Human Health-Current Prospective and Applications. *Front. Microbiol.* **2024**, *15*, 1487641. https://doi.org/10.3389/fmicb.2024.1487641.
4. Clarke, S. F.; Murphy, E. F.; O'Sullivan, O.; Lucey, A. J.; Humphreys, M.; Hogan, A.; Hayes, P.; O'Reilly, M.; Jeffery, I. B.; Wood-Martin, R.; Kerins, D. M.; Quigley, E.; Ross, R. P.; O'Toole, P. W.; Molloy, M. G.; Falvey, E.; Shanahan, F.; Cotter, P. D. Exercise and Associated Dietary Extremes Impact on Gut Microbial Diversity. *Gut* **2014**, *63* (12), 1913–1920. https://doi.org/10.1136/gutjnl-2013-306541.
5. Bein, A.; Shin, W.; Jalili-Firoozinezhad, S.; Park, M. H.; Sontheimer-Phelps, A.; Tovaglieri, A.; Chalkiadaki, A.; Kim, H. J.; Ingber, D. E. Microfluidic Organ-on-a-Chip Models of Human Intestine. *Cell. Mol. Gastroenterol. Hepatol.* **2018**, *5* (4), 659–668. https://doi.org/10.1016/j.jcmgh.2017.12.010.
6. Xu, Y.; He, B. The Gut-Muscle Axis: A Comprehensive Review of the Interplay between Physical Activity and Gut Microbiota in the Prevention and Treatment of Muscle Wasting Disorders. *Front. Microbiol.* **2025**, *16*, 1695448. https://doi.org/10.3389/fmicb.2025.1695448.
7. McKinney, J.; Velghe, J.; Fee, J.; Isserow, S.; Drezner, J. A. Defining Athletes and Exercisers. *Am. J. Cardiol.* **2019**, *123* (3), 532–535. https://doi.org/10.1016/j.amjcard.2018.11.001.
8. Di Vincenzo, F.; Del Gaudio, A.; Petito, V.; Lopetuso, L. R.; Scaldaferri, F. Gut Microbiota, Intestinal Permeability, and Systemic Inflammation: A Narrative Review. *Intern. Emerg. Med.* **2024**, *19* (2), 275–293. https://doi.org/10.1007/s11739-023-03374-w.
9. Longoria, C. R.; Guers, J. J.; Campbell, S. C. The Interplay between Cardiovascular Disease, Exercise, and the Gut Microbiome. *Rev. Cardiovasc. Med.* **2022**, *23* (11), 365. https://doi.org/10.31083/j.rcm2311365.
10. Scheiman, J.; Luber, J. M.; Chavkin, T. A.; MacDonald, T.; Tung, A.; Pham, L.-D.; Wibowo, M. C.; Wurth, R. C.; Punthambaker, S.; Tierney, B. T.; Yang, Z.; Hattab, M. W.; Avila-Pacheco, J.; Clish, C. B.; Lessard, S.; Church, G. M.; Kostic, A. D. Meta'omic Analysis of Elite Athletes Identifies a Performance-Enhancing Microbe That Functions via Lactate Metabolism. *Nat. Med.* **2019**, *25* (7), 1104–1109. https://doi.org/10.1038/s41591-019-0485-4.
11. Penggalih, M. H. S. T.; Sutanto, Y. S.; Taslim, N. A.; Syahputra, R. A.; Hardinsyah, H.; Tjandrawinata, R. R.; Nurkolis, F. Precision Nutrition in Sports Science: An Opinion on Omics-Based Personalization and Athletic Outcomes. *Front. Nutr.* **2025**, *12*, 1611440. https://doi.org/10.3389/fnut.2025.1611440.
12. Kuziel, G. A.; Rakoff-Nahoum, S. The Gut Microbiome. *Curr. Biol. CB* **2022**, *32* (6), R257–R264. https://doi.org/10.1016/j.cub.2022.02.023.
13. Rinninella, E.; Raoul, P.; Cintoni, M.; Franceschi, F.; Miggiano, G. A. D.; Gasbarrini, A.; Mele, M. C. What Is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* **2019**, *7* (1), 14. https://doi.org/10.3390/microorganisms7010014.
14. Magne, F.; Gotteland, M.; Gauthier, L.; Zazueta, A.; Pesoa, S.; Navarrete, P.; Balamurugan, R. The Firmicutes/Bacteroidetes Ratio: A Relevant Marker of Gut Dysbiosis in Obese Patients? *Nutrients* **2020**, *12* (5), 1474. https://doi.org/10.3390/nu12051474.
15. O'Brien, M. T.; O'Sullivan, O.; Claesson, M. J.; Cotter, P. D. The Athlete Gut Microbiome and Its Relevance to Health and Performance: A Review.Sports Med. Auckland. Nz **2022**, *52* (Suppl 1), 119–128. https://doi.org/10.1007/s40279-022-01785-x.
16. Kim, H. J.; Li, H.; Collins, J. J.; Ingber, D. E. Contributions of Microbiome and Mechanical Deformation to Intestinal Bacterial Overgrowth and Inflammation in a Human Gut-on-a-Chip. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (1), E7-15. https://doi.org/10.1073/pnas.1522193112.
17. Oertli, M.; Sundquist, M.; Hitzler, I.; Engler, D. B.; Arnold, I. C.; Reuter, S.; Maxeiner, J.; Hansson, M.; Taube, C.; Quiding-Järbrink, M.; Müller, A. DC-Derived IL-18 Drives Treg Differentiation, Murine Helicobacter Pylori-Specific Immune Tolerance, and Asthma Protection. *J. Clin. Invest.* **2012**, *122* (3), 1082–1096. https://doi.org/10.1172/JCI61029.
18. Schirmer, M.; Smeekens, S. P.; Vlamakis, H.; Jaeger, M.; Oosting, M.; Franzosa, E. A.; Horst, R. T.; Jansen, T.; Jacobs, L.; Bonder, M. J.; Kurilshikov, A.; Fu, J.; Joosten, L. A. B.; Zhernakova, A.; Huttenhower, C.; Wijmenga, C.; Netea, M. G.; Xavier, R. J. Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* **2016**, *167* (7), 1897. https://doi.org/10.1016/j.cell.2016.11.046.
19. Taverniti, V.; Minuzzo, M.; Arioli, S.; Junttila, I.; Hämäläinen, S.; Turpeinen, H.; Mora, D.; Karp, M.; Pesu, M.; Guglielmetti, S. *In Vitro* Functional and Immunomodulatory Properties of the Lactobacillus Helveticus MIMLh5-Streptococcus Salivarius ST3 Association That Are Relevant to the Development of a Pharyngeal Probiotic Product. *Appl. Environ. Microbiol.* **2012**, *78* (12), 4209–4216. https://doi.org/10.1128/AEM.00325-12.
20. West, N. P.; Pyne, D. B.; Cripps, A. W.; Hopkins, W. G.; Eskesen, D. C.; Jairath, A.; Christophersen, C. T.; Conlon, M. A.; Fricker, P. A. Lactobacillus Fermentum (PCC®) Supplementation and Gastrointestinal and Respiratory-Tract Illness Symptoms: A Randomized Control Trial in Athletes. *Nutr. J.* **2011**, *10*, 30. https://doi.org/10.1186/1475-2891-10-30.
21. Lahiri, S.; Kim, H.; Garcia-Perez, I.; Reza, M. M.; Martin, K. A.; Kundu, P.; Cox, L. M.; Selkrig, J.; Posma, J. M.; Zhang, H.; Padmanabhan, P.; Moret, C.; Gulyás, B.; Blaser, M. J.; Auwerx, J.; Holmes, E.; Nicholson, J.; Wahli, W.; Pettersson, S. The Gut Mi-

crobiota Influences Skeletal Muscle Mass and Function in Mice. *Sci. Transl. Med.* **2019**, *11* (502), eaan5662. https://doi.org/10.1126/scitranslmed.aan5662.
22. Sanders, K. M.; Koh, S. D.; Ro, S.; Ward, S. M. Regulation of Gastrointestinal Motility--Insights from Smooth Muscle Biology. *Nat. Rev. Gastroenterol. Hepatol.* **2012**, *9* (11), 633–645. https://doi.org/10.1038/nrgastro.2012.168.
23. Salarkia, N.; Ghadamli, L.; Zaeri, F.; Sabaghian Rad, L. Effects of Probiotic Yogurt on Performance, Respiratory and Digestive Systems of Young Adult Female Endurance Swimmers: A Randomized Controlled Trial. *Med. J. Islam. Repub. Iran* **2013**, *27* (3), 141–146.
24. Smarkusz-Zarzecka, J.; Ostrowska, L.; Leszczyńska, J.; Cwalina, U. Effect of a Multi-Strain Probiotic Supplement on Gastrointestinal Symptoms and Serum Biochemical Parameters of Long-Distance Runners: A Randomized Controlled Trial. *Int. J. Environ. Res. Public. Health* **2022**, *19* (15), 9363. https://doi.org/10.3390/ijerph19159363.
25. Di Vincenzo, F.; Del Gaudio, A.; Petito, V.; Lopetuso, L. R.; Scaldaferri, F. Gut Microbiota, Intestinal Permeability, and Systemic Inflammation: A Narrative Review. *Intern. Emerg. Med.* **2024**, *19* (2), 275–293. https://doi.org/10.1007/s11739-023-03374-w.
26. Van, K.; Burns, J. L.; Monk, J. M. Effect of Short-Chain Fatty Acids on Inflammatory and Metabolic Function in an Obese Skeletal Muscle Cell Culture Model. *Nutrients* **2024**, *16* (4), 500. https://doi.org/10.3390/nu16040500.
27. Barton, W.; Penney, N. C.; Cronin, O.; Garcia-Perez, I.; Molloy, M. G.; Holmes, E.; Shanahan, F.; Cotter, P. D.; O'Sullivan, O. The Microbiome of Professional Athletes Differs from That of More Sedentary Subjects in Composition and Particularly at the Functional Metabolic Level. *Gut* **2018**, *67* (4), 625–633. https://doi.org/10.1136/gutjnl-2016-313627.
28. Chandel, N. S. Basics of Metabolic Reactions. *Cold Spring Harb. Perspect. Biol.* **2021**, *13* (8), a040527. https://doi.org/10.1101/cshperspect.a040527.
29. Frampton, J.; Murphy, K. G.; Frost, G.; Chambers, E. S. Short-Chain Fatty Acids as Potential Regulators of Skeletal Muscle Metabolism and Function. *Nat. Metab.* **2020**, *2* (9), 840–848. https://doi.org/10.1038/s42255-020-0188-7.
30. Hsu, Y. J.; Chiu, C. C.; Li, Y. P.; Huang, W. C.; Huang, Y. T.; Huang, C. C.; Chuang, H. L. Effect of Intestinal Microbiota on Exercise Performance in Mice. *J. Strength Cond. Res.* **2015**, *29* (2), 552–558. https://doi.org/10.1519/JSC.0000000000000644.
31. Mika, A.; Van Treuren, W.; González, A.; Herrera, J. J.; Knight, R.; Fleshner, M. Exercise Is More Effective at Altering Gut Microbial Composition and Producing Stable Changes in Lean Mass in Juvenile versus Adult Male F344 Rats. *PLOS ONE* **2015**, *10* (5), e0125889. https://doi.org/10.1371/journal.pone.0125889.
32. Mika, A.; Fleshner, M. Early-Life Exercise May Promote Lasting Brain and Metabolic Health through Gut Bacterial Metabolites. *Immunol. Cell Biol.* **2016**, *94* (2), 151–157. https://doi.org/10.1038/icb.2015.113.
33. Nogal, A.; Valdes, A. M.; Menni, C. The Role of Short-Chain Fatty Acids in the Interplay between Gut Microbiota and Diet in Cardio-Metabolic Health. *Gut Microbes* **2021**, *13* (1), 1–24. https://doi.org/10.1080/19490976.2021.1897212.
34. Dziewiecka, H.; Buttar, H. S.; Kasperska, A.; Ostapiuk-Karolczuk, J.; Domagalska, M.; Cichoń, J.; Skarpańska-Stejnborn, A. Physical Activity Induced Alterations of Gut Microbiota in Humans: A Systematic Review. *BMC Sports Sci. Med. Rehabil.* **2022**, *14* (1), 122. https://doi.org/10.1186/s13102-022-00513-2.
35. Bressa, C.; Bailén-Andrino, M.; Pérez-Santiago, J.; González-Soltero, R.; Pérez, M.; Montalvo-Lominchar, M. G.; Maté-Muñoz, J. L.; Domínguez, R.; Moreno, D.; Larrosa, M. Differences in Gut Microbiota Profile between Women with Active Lifestyle and Sedentary Women. *PLOS ONE* **2017**, *12* (2), e0171352. https://doi.org/10.1371/journal.pone.0171352.
36. Xiong, R.-G.; Zhou, D.-D.; Wu, S.-X.; Huang, S.-Y.; Saimaiti, A.; Yang, Z.-J.; Shang, A.; Zhao, C.-N.; Gan, R.-Y.; Li, H.-B. Health Benefits and Side Effects of Short-Chain Fatty Acids. *Foods* **2022**, *11* (18), 2863. https://doi.org/10.3390/foods11182863.
37. Pugh, J. N.; Sparks, A. S.; Doran, D. A.; Fleming, S. C.; Langan-Evans, C.; Kirk, B.; Fearn, R.; Morton, J. P.; Close, G. L. Four Weeks of Probiotic Supplementation Reduces GI Symptoms during a Marathon Race. *Eur. J. Appl. Physiol.* **2019**, *119* (7), 1491–1501. https://doi.org/10.1007/s00421-019-04136-3.
38. Rowland, I.; Capurso, L.; Collins, K.; Cummings, J.; Delzenne, N.; Goulet, O.; Guarner, F.; Marteau, P.; Meier, R. Current Level of Consensus on Probiotic Science--Report of an Expert Meeting--London, 23 November 2009. *Gut Microbes* **2010**, *1* (6), 436–439. https://doi.org/10.4161/gmic.1.6.13610.
39. Kekkonen, R. A.; Vasankari, T. J.; Vuorimaa, T.; Haahtela, T.; Julkunen, I.; Korpela, R. The Effect of Probiotics on Respiratory Infections and Gastrointestinal Symptoms during Training in Marathon Runners. *Int. J. Sport Nutr. Exerc. Metab.* **2007**, *17* (4), 352–363. https://doi.org/10.1123/ijsnem.17.4.352.
40. Shing, C. M.; Peake, J. M.; Lim, C. L.; Briskey, D.; Walsh, N. P.; Fortes, M. B.; Ahuja, K. D. K.; Vitetta, L. Effects of Probiotics Supplementation on Gastrointestinal Permeability, Inflammation and Exercise Performance in the Heat. *Eur. J. Appl. Physiol.* **2014**, *114* (1), 93–103. https://doi.org/10.1007/s00421-013-2748-y.
41. Li, Z.; Li, Y.; Wang, Y.; Chen, J.; Liu, Y. The Athlete Gut Microbiome: A Narrative Review of Multi-Omics Insights and Next-Generation Probiotic Strategies. *Nutrients* **2025**, *17* (20), 3260. https://doi.org/10.3390/nu17203260.
42. Steiger, C.; Abramson, A.; Nadeau, P.; Chandrakasan, A. P.; Langer, R.; Traverso, G. Ingestible Electronics for Diagnostics and Therapy. *Nat. Rev. Mater.* **2019**, *4* (2), 83–98. https://doi.org/10.1038/s41578-018-0070-3.
43. Schork, N. J. Personalized Medicine: Time for One-Person Trials. *Nature* **2015**, *520* (7549), 609–611. https://doi.org/10.1038/520609a.
44. Jin, D.-M.; Morton, J. T.; Bonneau, R. Meta-Analysis of the Human Gut Microbiome Uncovers Shared and Distinct Microbial Signatures between Diseases. *mSystems* **2024**, *9* (8), e0029524. https://doi.org/10.1128/msystems.00295-24.
45. Su, G. L.; Ko, C. W.; Bercik, P.; Falck-Ytter, Y.; Sultan, S.; Weizman, A. V.; Morgan, R. L. AGA Clinical Practice Guidelines on the Role of Probiotics in the Management of Gastrointestinal Disorders. *Gastroenterology* **2020**, *159* (2), 697–705. https://doi.org/10.1053/j.gastro.2020.05.059.
46. O'Brien, M. T.; O'Sullivan, O.; Claesson, M. J.; Cotter, P. D. The Athlete Gut Microbiome and Its Relevance to Health and Performance: A Review. *Sports Med.* **2022**, *52* (Suppl 1), 119–128. https://doi.org/10.1007/s40279-022-01785-x.

Author

Megan Pan is a senior at Mountain View High School in California with a passion for medicine and sports. She plans on pursuing a pre-med track in the future.

DOI: 10.36838/IJHSR819.9

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

BioFETs with Magnetic Microbeads for Enhanced Biosensing and Early Disease Detection

Melody Wang[1], Feng Gao[2], Hsueh-Chia Chang[2]

1) BASIS Independent Fremont High School, Fremont, CA 94539, USA; melody.wang.world@gmail.com
2) Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA

ABSTRACT: Biological field-effect transistors (Bio-FETs) have demonstrated strong potential for biosensing and early disease detection with the new advances in nanomaterials, wafer fabrication, and integrated microfluidic devices. However, the intrinsic challenges, such as Debye length screening, electrical impedance interference, and biofouling effects from complex biological samples still exist in conventional Bio-FETs. Among emerging strategies, magnetic microbeads offer distinct advantages for enhancing sensing performance. Magnetic microbeads provide a high surface-to-volume ratio, enabling efficient conjugation and active capture of target biomolecules. Magnetic microbeads possess unique magnetic properties to allow precise manipulation by magnetic fields, such as controlled concentration, positioning, and mixing. Conjugated magnetic microbeads can be detected by both Bio-FETs and giant magnetoresistance (GMR) sensors, with improved robustness and sensitivity. This review focuses on Bio-FETs with magnetic microbeads, highlights key concepts, device types, manipulation, and signal detection mechanisms, and discusses remaining challenges and future opportunities for rapid, sensitive, portable, and wearable systems in precision healthcare.

KEYWORDS: Bio-FETs, Magnetic Microbeads, Conjugation, Microcoils, Detection.

Introduction

Over the past several decades, the semiconductor industry has undergone continuous innovations driven by Moore's law, leading to modern metal-oxide-semiconductor field-effect transistors (MOSFETs) with critical dimensions approaching the sub-2 nm node technology. As noted in the International Roadmap for Devices and Systems, advanced logic technology nodes have reached gate lengths and feature sizes on the order of a few nanometers, where atomic-scale effects increasingly dominate device behavior.[1] At these dimensions, transistor structures become comparable in size to fundamental biological molecules. For example, the DNA double helix has a uniform diameter of approximately 2 nm,[2] while single-stranded RNA molecules typically have an effective diameter of about 1nm.[3] Proteins occupy an intermediate nanoscale regime, as most globular proteins have diameters ranging from about 2 to 10 nm,[4] with antibodies measuring approximately 10-15 nm in overall dimension.[5] In contrast, viruses are larger nanoscale assemblies; they range in size from approximately 20-300 nm in diameter.[6]

As MOSFET dimensions are now comparable to biological entities, they are capable of detecting individual biomolecular or cellular components rather than bulk-averaged signals from large populations. Advanced 1D and 2D MOSFET layouts enable the fabrication of dense sensor lattice arrays, making simultaneous, massively parallel detection of millions of targets on a single chip feasible. Most importantly, because MOSFET operation is governed by electric-field changes of channel conductivity, these devices are intrinsically sensitive to nanoscale electrostatic and electromagnetic signals, enabling ultrafast detection on ps-ns timescales. Together, MOSFET dimensional matching, scalable array integration, and ultrafast electrical response have driven the rapid adoption of specialized biological field-effect transistors (Bio-FETs) for biological and biomedical applications.

Bio-FETs are biosensors that utilize field-effect transistors to convert biological recognition interactions directly into electrical signals. The fundamental operating principle was first introduced by Bergveld, who pointed out that changes in the ion concentration at the gate surface cause corresponding shifts of the transistor channel current.[7] In Bio-FETs, the traditional metal gate is replaced or functionalized with biological recognition elements such as DNA probes, antibodies, enzymes, or aptamers. When target biomolecules bind to the sensing surface, such as DNA hybridization, protein binding, or antigen-antibody interactions, their intrinsic charges change the surface potential, leading to a measurable shift in threshold voltage or channel conductance. Because Bio-FETs can operate at the nanoscale and measure the intrinsic electrostatic interactions without optical or chemical labels, they are particularly attractive for label-free and real-time detection.[8]

Despite significant advantages, Bio-FET sensors face several fundamental and practical limitations. One of the most critical challenges is electrostatic screening in physiological environments, commonly described by the Debye length. The Debye screening length in high-ionic-strength solutions can be reduced to less than 1 nm, severely limiting the sensitivity of charge-based biosensors.[9] This effect makes it difficult for Bio-FETs to detect biomolecules or binding events occurring several nanometers away from the sensing surface, such as antibody-antigen complexes. In addition, device-to-device variation, instability of surface functionalization, biofouling, and long-term signal drift are key engineering challenges. While Bio-FETs demonstrate exceptional sensitivity under

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.10

controlled conditions, reproducibility and robustness in real biological samples remain major obstacles.[10]

Therefore, future developments in Bio-FETs research have increasingly focused on engineering strategies that overcome Debye screening and improve robustness and reproducibility, including nanostructured and nanoporous channels,[11] reduced ionic strength, and signal-amplification approaches. Among these, the integration of magnetic microbeads has emerged as a particularly promising direction.[12] With a high surface-to-volume ratio for bioreceptor immobilization and active manipulation of target biomolecules using external magnetic fields, magnetic microbeads effectively increase the amount of molecular charges to the gate surface with pre-concentration and reduced background noise. Moreover, in addition to charge-based detection alone, magnetic microbead-based Bio-FETs also support hybrid sensing schemes, where electrical readout is combined with magnetic sensors, enhancing detection reliability in complex samples. Consequently, magnetic microbead-enhanced Bio-FET development represents a key future trend from high-sensitivity laboratory testing toward robust, scalable, and clinically capable biosensing platforms for early disease detection.

This review, therefore, focuses on the integration of magnetic microbeads with Bio-FET platforms, citing the capture of extracellular vesicles and exosomes for cancer biomarker detection as a clinical application for a promising noninvasive liquid biopsy strategy for early and sensitive diagnosis. The Bio-FET mechanism, different types, and the Debye length concept are introduced to establish the foundational understanding of Bio-FET systems. The review then examines microbead-based biosensors, with particular emphasis on available functionalization strategies and magnetic microbeads as mobile carriers and mobile nets for selective target conjugation and capture. Magnetic microbeads can be actively manipulated, especially by using microfabricated platforms, including planar microcoils, PCB-based coils, and 3D micro-solenoid structures, to achieve precise control at the targeted sensing interface. The increased charge concentration near the Bio-FET gate surface helps to overcome the limitations caused by Debye screening and poor mass transport.

The review compares the major post-manipulation detection strategies that convert bead accumulation into measurable outputs, including optical, electrochemical, electrical, and magnetic detection. Optical methods provide simple visual confirmation but are mostly qualitative. Electrochemical techniques such as cyclic voltammetry and impedance spectroscopy offer high sensitivity for clinical biomarker assays, but they often require redox-active labels, enzymes, or additional chemical modification steps rather than being fully label-free. They also suffer from nonspecific adsorption and unwanted side reactions from other molecules present in complex biological samples, as well as electrode fouling. Magnetic detection using giant magnetoresistance (GMR) and tunnel magnetoresistance (TMR) sensors enables highly quantitative bead counting without being affected by Debye screening; however, the major limitation is the complexity of device fabrication and integration with highly specialized magnetic sensor structures. Overall, Electrical detection using Bio-FETs and magnetic microbeads enables ultrasensitive detection of targets such as exosomal miRNA and DNA through charge-based sensing, creating a highly promising platform for early disease detection by reducing sample pretreatment, minimizing false signals, and enabling direct analysis in minimally processed biological samples.

Discussion

Clinical Aspects of the Applications:

Cancer remains one of the leading causes of death worldwide, with approximately 20 million new cases and 9.7 million deaths reported globally in 2022.[13] Lung, breast, and colorectal cancers are among the most common and deadliest types. However, patient survival rates are strongly dependent on the stage at which the disease is diagnosed. Early-stage detection results in survival rates as high as 90% for breast and ovarian cancer and approximately 80% for lung cancer, while survival rates drop sharply to 15% or lower when the disease is diagnosed at advanced stages. This huge difference highlights the urgent need for early, sensitive, noninvasive, and point-of-care diagnostic methods. Improving technologies for early disease detection, particularly through liquid biopsy approaches that can identify biomarkers before symptoms appear, has the potential to lower cancer-related deaths and help patients receive treatment earlier and more effectively.

Extracellular vesicles (EVs), including exosomes, are nanoscale extracellular vesicles found in bodily fluids such as blood, saliva, urine, and breast milk. EVs contain proteins and RNA that represent the biological state of the parent cells, making them unique carriers of disease-related biomarkers.[14] By isolating and analyzing these molecular contents, EV-based liquid biopsy approaches can support early cancer detection, monitor treatment response, and help predict disease progression, enabling noninvasive analysis and personalized healthcare.

Despite significant progress, EVs/Exosomes detection using Bio-FET sensors has encountered the electrostatic screening of the nanometer-level Debye length.[15] Magnetic microbeads functionalized with capture probes have been proven to effectively enable selective binding and active manipulation of EVs as an interim medium, and for low-concentration detection effectively.[16,17] Solovicova *et al.* provided a comprehensive overview of magnetic-bio-affinity capture applied to the detection of exosomes or exosomal cargo as disease markers using different affinity capture ligands.[18] Wu *et al.* developed the Mag-Net platform, which uses magnetic beads to capture and enrich EVs from plasma while depleting background proteins, thereby enabling sensitive downstream detection of EV-associated proteins in low-concentration clinical samples.[19]

Fundamentals of Bio-FET and Bio-FET Types:

A Bio-FET is constructed by integrating biological detection with an FET platform. The fundamental structure of a Bio-FET consists of the following key components: 1) Semiconductor Channel as the conduction pathway between the source and drain electrodes. 2) Gate Dielectric Layer as both

an insulating layer and a sensing interface. 3) Source and Drain Electrodes and 4) Functionalized Sensing Surface as shown in Figure 1.

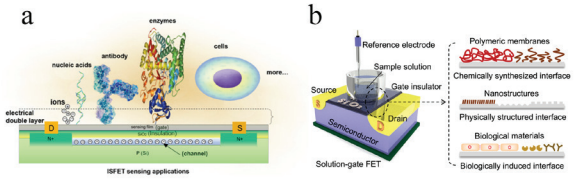


Figure 1: (a) Schematic illustration of the BioFET-sensing applications for different analytes. Reprinted with permission from Electrochemical Science Advances.[20] (b) Conceptual structure of potentiometric FET biosensor, consisting of a biological sample, a signal transduction interface, and a detection device. Reprinted with permission from Commun. Chem.[21]

Pandey *et al.* provided a comprehensive and up-to-date review of bio-field-effect transistors (Bio-FETs), systematically covering the fundamental operating principles of FETs, as well as the structural design and fabrication strategies specific to Bio-FET devices.[22] The review categorizes and analyzes major Bio-FET types based on their applications, including DNA-FETs, enzyme-FETs (ENFETs), immuno-FETs, ion-sensitive FETs (ISFETs), and aptamer-based FETs, supported by detailed comparisons of different case studies for each type. In addition, the paper provides an in-depth discussion of the most widely reported materials in Bio-FET development, such as laser-induced graphene, carbon nanotubes, reduced graphene oxide, and molybdenum disulfide (MoS_2).

Instead of categorizing Bio-FETs by applications and detection materials, this review will follow Rai *et al.* by structuring the analysis around the various FET architectures and their performances, as shown in Figure 2, including MOSFET, ISFET (Ion Sensitive Field Effect Transistor), FGFET (Floating Gate Field Effect Transistor), EGFET (Extended Gate Field Effect Transistor), Carbon-Based FETs (Carbon Nanotube FET and Graphene FET), TFET (Tunnel Field Effect Transistor).[23]

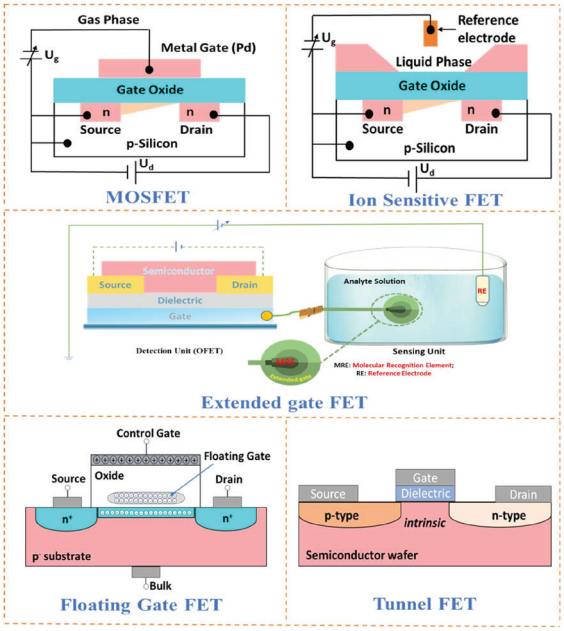


Figure 2: Various architectures of FET technology. Reprinted with permission from Talanta Open.[23]

In biosensing applications, MOSFET-based sensors allow small changes induced by biomolecular interactions to produce measurable current variations directly. Sadighbayan *et al.* developed a commercial MOSFET-based biosensor to detect C-reactive protein (CRP).[24] The device identifies CRP by measuring changes in the drain current that occur when CRP binds to its specific antibody. Jieun Lee *et al.* reported a silver nanowire used to detect the target in Figure 3a, and a hybrid MOSFET converted this signal into an electrical output.[25] Both parts were integrated on a single chip using conventional CMOS fabrication.

ISFETs (Ion Sensitive Field Effect Transistors) are similar to MOSFETs, except that the metal gate is replaced by a reference electrode or sensitive membrane deposited on top of the gate, which is used to provide bias voltage (Vg) in an ISFET. Manimekala *et al.* covered a wide range of ISFETs on Dopamine, Glucose (non-enzymic), Nucleic Acid (DNA), Protein, Virus, and Cancer Biomarkers detection using nanomaterials, especially 1D nanowires and nanoparticles, as well as 2D graphene in Figure 3b.[26]

FGFETs (Floating Gate Field Effect Transistors) physically isolate the transistor channel from the electrolyte with improved device robustness. Panahi *et al.* have successfully utilized FGFETs for pH sensing, enzymatic reaction monitoring, and bioelectronic recording in Figure 3c.[27]

EGFETs (Extended Gate Field Effect Transistors) are standard MOSFETs, with the gate extended to the reference electrode, electrolyte, and sensing layer, making them disposable sensor heads, with ease of packaging, resistance to light and temperature variations, and straightforward replacement of sensing films. Liu *et al.* presented a molecularly imprinted polymer (MIP)-based EGFET sensing system that is used for the measurement of CRP concentration in human serum.[28] Janićijević *et al.* illustrated the evolution of EGFETs in Figure 3d as an advanced class of FET biosensors that balance the high sensitivity of field-effect transduction with practical benefits.[29]

Carbon-Based FETs, including CNTFETs and GFETs, have emerged as powerful biosensing platforms due to their exceptional electrical, mechanical, and surface properties. Yixi Deng *et al.* focus on applying CNTFET biochemical sensors to detect biomarkers, as shown in Figure 4a.[30] The functionalization step is key to making CNTs interact specifically with target analytes (e.g., DNA, proteins). Jiahao Li *et al.* demonstrated an AuNP-decorated GFET nanosensor for rapid and amplification-free SARS-CoV-2 identification, as shown in Figure 4b.[31] Importantly, the nanosensor exhibited strong anti-interference capability and promising clinical applicability.

TFETs (Tunnel Field Effect Transistors) are novel biosensors based on a complementary metal oxide semiconductor (CMOS)-compatible silicon nanowire tunneling field-effect transistor (SiNW-TFET).[32] A TFET works by band-to-band tunneling instead of electrons going over a barrier. The mechanism is illustrated in Figure 4c. Gao *et al.* fabricated a "top-down" TFET with a low-cost anisotropic self-stop etching technique.[33] Real-time measurements highlighted the ability to use clinically relevant samples such as serum. A

DOI: 10.36838/IJHSR819.10

TFET is a special type of transistor designed to use much less power than conventional transistors. Like other transistors, it has a source, drain, and gate, but it works in a different way, as shown in Figure 4d. Ongoing improvements in device structure and materials are making them promising candidates for next-generation, energy-efficient biosensors.[34,35]

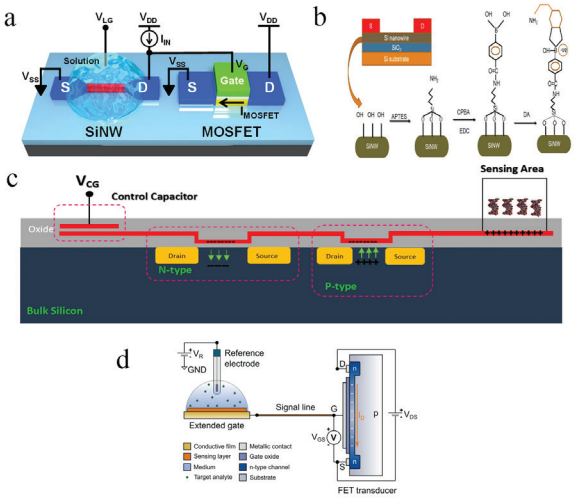


Figure 3: Fundamental Bio-FET types: (a) Schematic diagram of the demonstrated biosensor composed of an n-type SiNW and a MOSFET. Reprinted with permission from Sci. Rep.[25] (b) ISFET - A SiNW Bio-FET for DA detection. Reprinted with permission from J. Electron Mater.[26] (c) The FGFET structure and a real chip designed for Label-free DNA analysis. Reprinted with permission from Biosensors.[27] (d) Illustration of the EG-FET sensing concept using the example of an n-type enhancement mode MOSFET transducer. Reprinted with permission from Next Nanotechnology.[29]

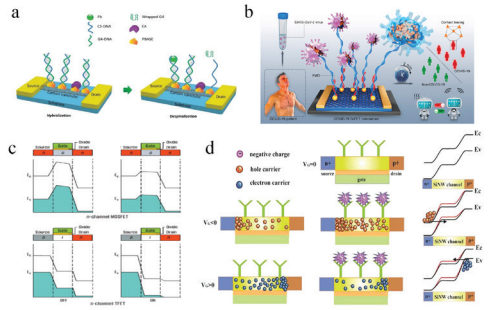


Figure 4: Advanced Bio-FET types: (a) Hybridization and desalination of CNT-FET sensors. Reprinted with permission from Biosensors.[30] (b) GFET nanosensor. Reprinted with permission from Biosens. Bioelectron.[31] Copyright 2021 Elsevier. (c) An n-channel MOSFET and an n-channel TFET turn on and off by showing their energy band diagrams. MOSFET ON: electrons go over the barrier. TFET ON: electrons tunnel through the barrier.[32] Copyright 2016 Elsevier. (d) SiNW-TFET mechanism. Reprinted with permission from Sci. Rep.[33]

A summary of different types of Bio-FETs and their applications is presented in Table 1. The table compares the various Bio-FET types, including their gate or interface structures, key sensing mechanisms, major advantages and limitations, and typical applications.

Table 1: Different Types of Bio-FETs and Their Applications.

FET Type	Gate / Interface Structure	Key Sensing Mechanism	Major Advantages	Main Limitations	Typical Applications
MOSFET [24, 25]	Metal gate + oxide (SiO_2)	Gate-voltage-modulated channel conductivity	High sensitivity, CMOS-compatible, low noise	Back-gate limited for neutral species	Protein-sensing CRP Hybrid CMOS biosensors
ISFET [26]	Ionic solution and membrane + reference electrode	Ion-induced surface potential modulation	Small size, low cost, scalable arrays	Reference electrode, drift	pH sensing, Enzymatic biosensors, SoC platforms
FGFET [27]	Floating sensing gate + control gate	Chemical-to-electrical signal conversion	No reference electrode, robust, CMOS-compatible	Structural complexity	pH, Enzyme sensing Bioelectronics arrays
EGFET [28,29]	Extended external gate with sensing film	Remote gate modulation	Disposable head, high stability, high sensitivity	Limited scaling, CMOS integration gap	pH, Enzyme detection Protein
Carbon-Based FET [30,31]	CNT or graphene channel	Charge transfer at the nanomaterial surface	Ultra-high sensitivity, fast response, low power	Fabrication variability	Gas sensing Virus Neurotransmitter
TFET [33]	p-i-n structure with tunneling	Band-to-band tunneling	Ultra-low power, steep sub-threshold swing (SS)	Low ON current, ambipolarity	Low-power biosensors

Debye Length:

The Debye length refers to the characteristic length scale over which electric potentials are screened by mobile ions in an electrolyte, and it essentially defines the thickness of the electrical double layer adjacent to a charged surface, as pointed out by H.-C. Chang *et al.*[36] The Debye length is the distance from the charged surface where the potential has decayed to the 1/e point or 36.7% of the surface potential. Physically, it characterizes the thickness of the diffuse part of the electrical double layer (EDL) formed when an electrolyte solution is in contact with a charged surface. The Debye length quantifies how quickly the electric field is screened by ions in solution, and it appears explicitly in the Poisson-Boltzmann model of the EDL - a foundational concept of double layers.

Debye length λ_D scales inversely with the square root of concentration and also scales inversely with the valence of the ions in solution for electrolytes. Therefore, the high ionic strength subsequently leads to a short Debye length; the Debye length λ_D for a divalent ion is half that for the monovalent ion. To put λ_D into perspective for operational systems, at an electrolyte concentration of 1 μM for a 1:1 electrolyte, λ_D=304 nm, but at 1 M, it is 0.3 nm. For typical biological Phosphate buffer solutions (PBS) containing a mixture of potassium monohydrogen phosphate (K2HPO4) and potassium dihydrogen phosphate (KH2PO4), in 1x PBS, the Debye length is typically ~0.7 nm or less than 1 nm, where most biomolecular charges are screened.[37] For 0.1x-1x PBS, reported Debye lengths range from ~0.7 nm up to ~2 nm as the buffer is diluted (lower ionic strength gives longer Debye length), as shown in Figure 5.[38]

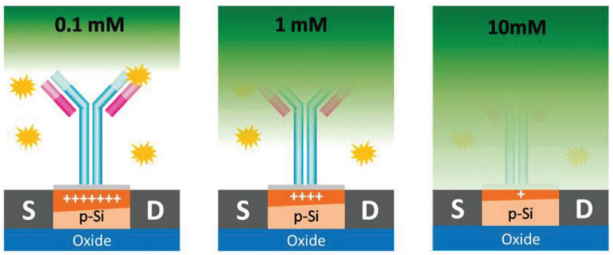


Figure 5: Debye Length Effects under different electrolyte concentrations. Reprinted with permission from J. Am. Chem. Soc.[38]

Zhang and Zheng discuss the reduced FET sensitivity using the example of the negatively-charged streptavidin molecules, which are surrounded by positively charged ions due to the electrostatic force.[39] Conventional FET sensors sense the concentration of the target species by their intrinsic charge. The charges of solution-based molecules, however, can be screened by dissolved counterions in the solution. Their work shows that the Debye length can be increased by using diluted, low-ionic-strength buffers without significantly affecting biomolecular interactions such as oligonucleotide hybridization; however, additional strategies to mitigate Debye screening are still required to improve FET sensitivity.

Microbeads and Magnetic Microbeads: Properties and Functions:

Microbead-based biosensors have emerged as a promising platform due to their versatility, strong signal responses, and ability to detect multiple targets simultaneously, offering several inherent advantages such as high surface-to-volume ratio, rapid response, high throughput, and minimal sample volume. Hui-Pin Cheng *et al.* compare the advantages and disadvantages of microbead-based biosensors in Table 2 and categorize different microbead-based biosensors by their different materials as shown in Figure 6, highlighting the magnetic beads for improving specificity and reducing background noise.[40]

Table 2: Advantages and disadvantages of microbead-based biosensors.[40]

Pros	Cons
High Surface to Volume Ratio High Throughput Fast Response Reduced Matrix Effects Enhanced Sensitivity and Dynamic Range Multiplexed Detection Minimal Sample Volume Miniaturization and Portability	Reproducibility and Standardization Cross-Reactivity Instrumentation Complexity Clinical Translation

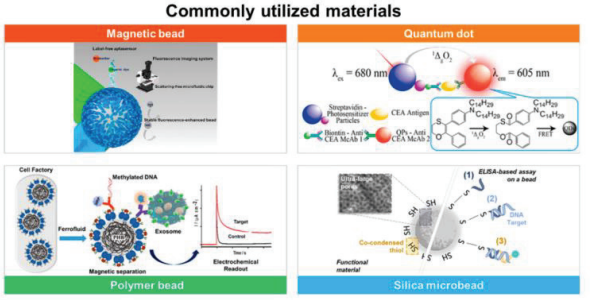


Figure 6: Commonly utilized materials in microbead-based biosensors. Reprinted with permission from Sensors.[40]

Magnetic microbeads, typically composed of iron oxide or iron oxide within embedded polymer matrices, can be precisely controlled to mitigate Debye Length screening effects. Under an applied magnetic field, magnetic microbeads undergo directed motion and mixing, generating localized fluid flows. Although the Debye length remains fixed by electrolyte ionic strength, these localized fluid flows enhance mass transport and reduce diffusion limitations. As a result, target molecules are delivered more efficiently to the BioFET sensing surface, improving effective signal response despite strong Debye screening.

In addition, magnetic microbeads have a high surface-to-volume ratio, which can be biofunctionalized with target molecules, and the conjugates can be concentrated using magnetic fields, increasing both the sensitivity and specificity of detection in complex biological samples. Magnetic microbeads benefit from the absence of magnetic background signals in biological samples, resulting in inherently low noise and high specificity. Therefore, unlike fixed sensing substrates, magnetic microbeads act as mobile carriers and mobile nets that transport captured biomolecules to defined sensing regions, diminishing the difficulties posed by the Debye length layer, mixing at low Reynolds numbers, or high flow resistances at small scales.

Furthermore, magnetic microbeads are commercially available in different sizes, magnetic properties, and surface coatings, making them very versatile, indispensable tools in biosensing, especially to realize integrated and miniaturized technologies based on small fluid volumes.

Magnetic Microbead Conjugation for Biosensing Applications:

Magnetic microbead conjugation is a key step that transforms simple manipulable particles into highly selective biosensing tools. By attaching antibodies, aptamers, peptides, or nucleic acids onto their surface, magnetic microbeads can specifically recognize and capture target biomolecules such as proteins, exosomes, and cancer biomarkers. This makes magnetic bead conjugation an important strategy for improving sensitivity and specificity in Bio-FETs and clinical diagnostic applications.

Modh *et al.* have reviewed magnetic microbeads in biosensing with different detection methods.[41] For selectivity, magnetic microbeads are decorated with different biological recognition elements, as shown in Figure 7a: (1) Antibodies target specific protein biomarkers or pathogens in immunoassays. (2) Aptamers - synthetic single-stranded DNA or RNA sequences selected by SELEX (Systematic Evolution of Ligands by EXponential enrichment) - offer advantages such as ease of chemical modification, high stability, and the ability to design unique sensing formats (e.g., structure switching, induced dissociation). (3) Peptides and other small ligands can also be used, depending on the target and application. Functionalization strategies typically rely on established bioconjugation chemistry, e.g., covalent linking through reactive surface groups, streptavidin-biotin interaction, or linker arms for aptamer immobilization.

Yue *et al.* have developed a label-free electrochemical biosensor based on magnetically induced self-assembly. $Mg0.5Cu0.5Fe_2O_4$-Au nanocomposites are prepared and assembled on a magnetic glassy carbon electrode (MGCE) under a magnetic field, forming the active sensor interface. A thiol-modified single-stranded DNA aptamer specific to CA125 is attached to the Au surface via Au-S bonds, without using a coupling agent.[42] The CA125 antigen, which is an important indicator for early detection of ovarian cancer, is captured directly by the aptamer through specific binding (aptamer - antigen interaction). After aptamer immobilization and antigen capture, BSA is used for surface blocking to reduce nonspecific adsorption.

DOI: 10.36838/IJHSR819.10

Xiaoxia Fang *et al.* have developed a magnetic bead selective adsorption strategy (MagExo) for easy-to-operate extracellular vesicles (EVs) isolation, as shown in Figure 7b and 7c.[43] EVs can be selectively adsorbed on the surface of magnetic microbeads in the presence of a hydrophilic polymer and can be separated from biological fluids with high purity by simple permanent magnetic separation. The MagExo procedure captured 2X higher yield and comparable purity of the harvested EVs to that obtained by ultracentrifugation (UC). The target RNA amount of the isolated EVs by MagExo was almost dozens and hundreds of times higher than the gold standard Density gradient ultracentrifugation (DG-UC) and ultracentrifugation (UC) methods, respectively. More detailed and latest magnetic affinity-based methods for isolating and detecting exosomes can be found in the review of Veronika Solovicová *et al.*, including antibody-based capture, DNA aptamer-based capture, peptides and glycan recognition, and other recognition schemes.[18]

Lin *et al.* addressed two main problems when using ISFETs to detect proteins.[44] First, antibodies on the sensor surface are often longer than the Debye length (~0.7 nm) in biological solutions, which makes it difficult for the bound proteins to be tightly attached to the sensor surface. Second, many proteins carry a small electrical charge and can be easily affected by the pH of the surrounding solution, leading to weak changes in surface potential and poor signal response. To solve these issues, the authors used a magnetic bead-based detection approach, as in Figure 7d, where an external magnetic field pulls the bead-protein complexes close to the sensor surface. After protein capture, DNA strands are attached to the microbeads as labels. DNA carries a high and well-defined negative charge, which is much easier for semiconductor sensors to detect than protein charges. Their work demonstrates a hybrid system combining magnetic microbeads, DNA-based signal amplification, and a semiconductor sensor within a microfluidic chip, enabling sensitive and fast protein biomarker detection while overcoming Debye screening limitations.

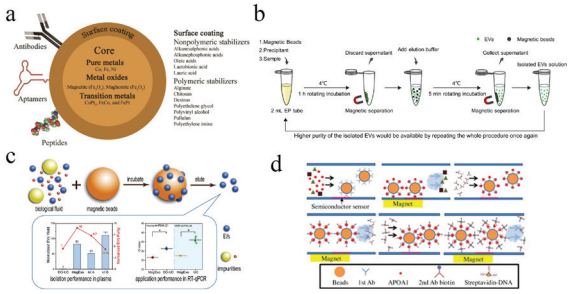


Figure 7: Magnetic microbead conjugation: (a) Composition of magnetic microbeads (MBs) used in analytical applications. Reprinted with permission from Sensors.[41] (b) and (c) Schematics of MagExo for EV extraction. Reprinted with permission from Acta Biomater.[43] Copyright 2021 Elsevier. (d) Magnetic microbeads conjugated with multiple antibodies. Reprinted with permission from Anal. Chim. Acta.[44] Copyright 2015 Elsevier.

Magnetic Microbead Manipulation in Microfabricated Systems:

Besides using permanent magnets, magnetic microbeads can be effectively integrated with microfabricated microcoil or micro-solenoid systems, which can be used to generate a localized magnetic field and magnetic field gradient to attract, trap, move, or concentrate magnetic microbeads near a sensing region (e.g., a Bio-FET gate or microfluidic channel). Microcoils can be applied with different magnetic actuation modes, including DC current for static bead trapping, AC current for bead agitation and mixing, and pulsed current for capture - release cycles. These actuation modes enable efficient bead concentration, washing steps, and controlled contact between magnetic beads and the sensing surface. This allows more precise bead manipulation and supports a wide range of magnetic biosensors based on magnetic, electrical, optical, and electrochemical detection methods.

Vania Silverio *et al.* have developed the manipulation of magnetic microbeads for therapeutic drug monitoring with thin film microelectromagnet traps, taking into consideration both magnetic trapping and integrated efficient cooling as in Figure 8.[45] Six different designs (thickness, width, number of turns, positioning of coils, etc.) of planar microelectromagnets traps (MEMT) with four external coils in series and one central coil connected for an opposite direction of manipulation of magnetic microbeads in microfluidic flows. The yield of local concentration was highly dependent on coil design, but the optimized design showed a 210% increase. Their work demonstrated planar microelectromagnet can be fabricated and utilized in magnetic microbead concentration and manipulation for higher-sensitivity detection.

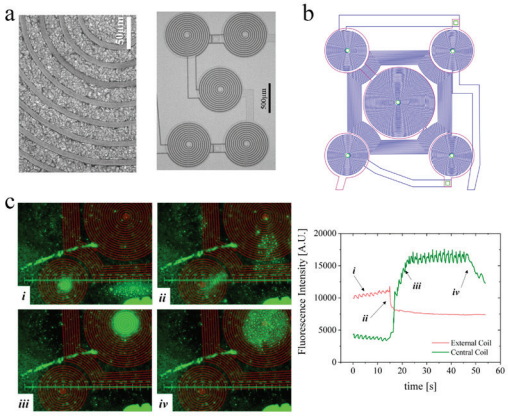


Figure 8: Magnetic microbead in microfabricated systems: (a) and (b) Microfabricated microcoil designs. (c) Experimental activation and deactivation pulses resulting in MB concentration. Reprinted with permission from Micromachines.[45]

Planar microfabricated copper coils can be produced using standard photolithography and electroplating processes. In addition, printed circuit board (PCB)-based fabrication offers a cost-effective and scalable alternative for implementing microcoils without the need for complex microfabrication steps. Roer Eka Pawinanto *et al.* have proposed using PCB technology to fabricate planar micro-coils for driving magnetic MEMS actuators, as shown in Figure 9a.[46]

3D (out-of-plane) microcoil or solenoid structures differ from planar spiral coils and are commonly employed when strong vertical magnetic fields and field gradients are required for magnetic bead manipulation near sensors. Heuberger *et*

al. developed a deep-UV lithography-based 3D microforming process capable of fabricating thick, high-aspect-ratio microstructures up to 200 μm thick with steep sidewalls in Figure 9b.[47] Hoa Thanh Le *et al.* have reviewed the advances in MEMS inductor fabrication, including Planar MEMS Inductors, 3D MEMS Inductors, and especially 3D air-core and magnetic-core designs as shown in Figure 9c.[48,49] The authors shared the discussions on how advances in microfabrication, materials, and 3D architectures have transformed inductors from bulky passive components into high-performance, chip-scale devices. A comparison of planar microcoil and 3D micro-solenoid is listed in Table 3.

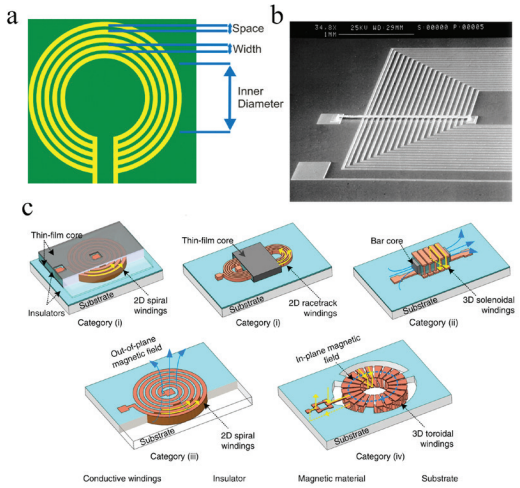


Figure 9: Microcoil designs (a) Microcoil for magnetic MEMS actuator by PCB technology.46 (b) Planar gold microcoil with integrated contact bridge with optical DUV-lithography for high microstructures.[47] (c) 3D illustrations of different categories of MEMS inductors. Reprinted with permission from Microsyst Nanoeng.[48,49]

Table 3: Comparison of Planar Microcoil and 3D Micro-solenoid:

Feature	Planar Coil	3D Solenoid
Coil geometry	Flat spiral (2D)	Vertical windings (3D)
Main field direction	In-plane	Out-of-plane (vertical)
Vertical bead force	Moderate	Strong
Bead localization	Broad area	Highly localized
Drive current	Higher	Lower
Fabrication complexity	Low	Medium - High
CMOS compatibility	Excellent	Partial/Hybrid
Best use case	Mixing, transport, and large-area trapping	Bead capture, pre-concentration, Bio-FET sensing

Magnetic Microbead Manipulation in Microfluidic Chips:

Precise magnetic microbead manipulation in microfluidic chips has also been widely studied, and these systems provide valuable design strategies for integrating microcoils with Bio-FET sensors.

Lefebvre, O. *et al.* have built a microfluidic chip with integrated planar microcoils that can effectively trap magnetic nanobeads (~300 nm) and also track temperature rise during operation (important because coil current causes Joule heating).[50] Their work includes design, fabrication, and operation for trapping magnetic nanobeads with embedded planar microcoils inside a microfluidic chip, providing a practical framework to optimize planar coil geometry (e.g., line width/spacing, inner/outer dimensions) rather than trial-and-error. Cai *et al.* provide a thorough, in-depth overview of how magnetic microbeads are manipulated within microfluidic chips and how these manipulation strategies are applied in biological and diagnostic workflows.[51] The review provides the physical basis for magnetic manipulation, including superparamagnetism and magnetic bead properties, magnetic forces and magnetophoresis, and bead surface chemistry and surface modification methods. The review systematically explains how bead manipulation techniques are used in real biological applications, such as biomarker and analyte capture, cell sorting, immunoassays and ELISA variants, and extraction/enrichment of nucleic acids or extracellular vesicles.

Detection Strategies for Manipulated Magnetic Microbeads:

After magnetic microbeads are concentrated, several practical detection strategies can convert the bead accumulation into measurable signals. These approaches include simple visual observation and advanced electrical and magnetic sensing techniques. In general, four main detection methods are used: optical detection, electrochemical detection, electrical detection using Bio-FETs, and magnetic detection based on giant or tunnel magnetoresistance sensors.

Optical Detection:

Optical detection is the simplest and most commonly used method to observe magnetic microbeads after they are concentrated by a microcoil. After bead concentration, the region near the coil or sensor surface can be examined using a standard optical microscope, or a fluorescence microscope if fluorescent microbeads are used. Bead accumulation is identified by an increase in bead density or changes in brightness or fluorescence intensity, as in Figure 10a.[41] This method is very easy to implement and provides direct visual confirmation; however, it is mostly qualitative and has limited sensitivity when detecting low numbers of microbeads. As a result, optical detection is mostly for proof-of-concept demonstrations and microfluidic experiments.

Electrochemical Detection:

Cyclic Voltammetry (CV) and Electrochemical Impedance Spectroscopy (EIS) are the main electrochemical techniques, as shown in Figure 10b. Magnetically induced self-assembly of functionalized microbeads enables highly sensitive electrochemical detection of cancer biomarkers in complex samples, where the measured current or impedance change is proportional to bead concentration.[52] This strategy offers high sensitivity and robustness in real biological matrices, since electrochemical methods are less affected by sample complexity. However, because the approach relies on enzyme or redox labeling, it is not fully label-free and requires additional chemical modification steps. Therefore, electrochemical bead-based detection is particularly well suited for clinical assays and protein biomarker detection, as discussed by Cheng, H.-P. *et al.*[40]

Electrical Detection for Bio-FETs with Magnetic Microbeads:

Magnetic bead-assisted Bio-FET detection represents a highly sensitive strategy due to its superparamagnetic prop-

DOI: 10.36838/IJHSR819.10

erties, a key advantage for mixing, isolating, or concentrating targets from complex samples. In the following three studies, magnetic microbeads demonstrated their abilities to effectively and selectively capture important disease biomarkers under various situations. External magnetic fields help pull these targets closer to the graphene FET sensing surface, which improves target delivery and makes the electrical signal stronger and easier to measure at extremely low levels like femtomolar and picomolar concentrations.

Yu *et al.* have developed an integrated microsystem that combines magnetic bead-based exosome isolation with a FET biosensor to enable selective capture and highly sensitive detection of tumor-derived exosomal microRNA (miRNA), specifically miRNA10b.[53] The system uses magnetic microbeads functionalized with two antibodies - GPC-1 and EpCAM - which bind to surface proteins highly expressed on pancreatic cancer-derived exosomes. These dual-antibody microbeads selectively recognize and isolate pancreatic cancer exosomes from plasma samples before sensing. After concentration, the isolated exosomes are processed to release their internal miRNA. The released miRNA10b is then detected using a Peptide Nucleic Acid (PNA)-functionalized reduced GFET biosensor. PNAs hybridize selectively with the target miRNA, enabling highly specific and sensitive electrical detection down to an ultrasensitive detection limit of ~78 fM for miRNA10b. The method was validated with 40 clinical plasma samples, effectively distinguishing pancreatic cancer patients from healthy controls. The detailed process is illustrated in Figure 10c.

Sun *et al.* developed a GFET by transferring a chemical-vapor-deposited graphene film onto a substrate to serve as the conductive channel, as shown in Figure 10d.[54] Graphene's high surface area and excellent electronic properties make it an attractive material for sensitive FET biosensing. To enable specific DNA detection, the graphene surface was chemically modified with 1-pyrenebutanoic acid succinimidyl ester (PBASE), which acted as an anchoring layer for immobilizing a probe aptamer. The aptamer selectively binds complementary single-stranded DNA targets. Target complementary DNA strands were magnetically labeled using magnetic nanobeads. These magnetically tagged DNA targets could be manipulated by applying an external periodic magnetic field, causing them to interact dynamically with the graphene surface. Under the periodic magnetic field, the impedance of the GFET exhibited periodic oscillations, and the amplitude of these impedance changes correlated with DNA concentration. The biosensor demonstrated high sensitivity, with a detection limit on the order of ~1 pM for single-stranded DNA.

Zhu *et al.* developed a magnetically controlled GFET biosensor for highly sensitive detection of cardiac troponin I (cTnI), an important biomarker for acute myocardial infarction.[55] In this system, graphene films transferred onto Indium Tin Oxide (ITO) conductive glass served as the conductive channel, while a cTnI-specific aptamer was immobilized on the graphene surface using PBASE as the linker for selective target capture. Magnetic nanobeads functionalized with cTnI antibodies were then introduced to form an aptamer/cTnI/antibody/magnetic bead sandwich structure. By applying an external magnetic field, the magnetic beads were actively pulled closer to the graphene sensing surface, significantly enhancing the electrical signal and improving detection sensitivity to a low detection limit of 10 pM for cTnI.

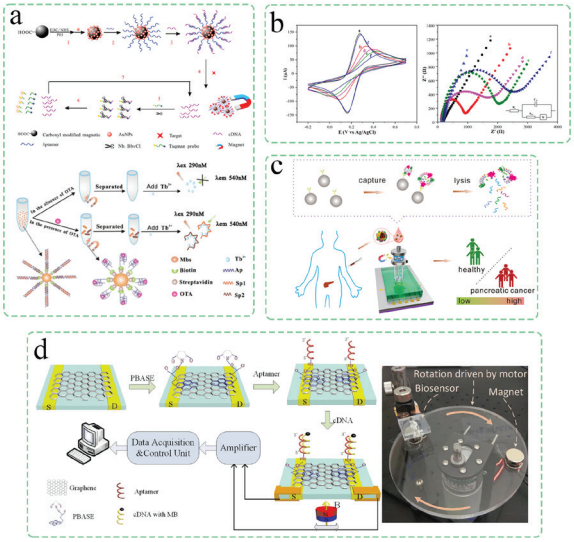


Figure 10: Different detection strategies for manipulated magnetic microbeads: (a) Fluorescence-based assays. Reprinted with permission from Sensors[41] (b) CV profiles for Electrochemical Detection. Reprinted with permission from Arab J. Chem.[52] (c) Integrated FET sensing microsystem for specific detection of pancreatic cancer exosomal miRNA10b. Reprinted with permission from Anal. Chim. Acta.[53] Copyright 2023 Elsevier. (d) Magnetic Graphene Field-Effect Transistor Biosensor for Single-Strand DNA Detection. Reprinted with permission from Nanoscale Res. [54]

The research gap is that most studies either focus on magnetic bead-assisted Bio-FET sensing or microcoil-based bead manipulation separately, while very few directly integrate on-chip microelectromagnets with Bio-FET detection on the same platform. This makes microcoil-driven magnetic bead Bio-FET systems a strong and still novel research direction for early disease detection. Preliminary work at Chang's Lab has demonstrated the feasibility of this concept. A microcoil-driven magnetic bead Bio-FET system can be described as follows: Magnetic microbeads are introduced into the biological sample with conjugation agents and act as magnetic carriers to capture target biomolecules. When current is applied to the on-chip microcoil, the generated localized magnetic field and field gradient actively pull the magnetic microbeads toward the Bio-FET sensing surface. After the beads are concentrated near the gate region, only the correctly conjugated bead-target complexes specifically bind ("dock") to the functionalized Bio-FET surface, while non-target or non-conjugated beads are released once the magnetic field is reduced. The retained magnetic microbeads change the local charge environment at the Bio-FET surface, resulting in a stronger and more measurable electrical response through changes in gate voltage or channel current for Bio-FET detections.

Magnetic Detection:

Magnetic detection using giant magnetoresistance (GMR) or tunnel magnetoresistance (TMR) sensors provides a direct and quantitative way to detect magnetic microbeads concentrated by a microcoil. In this approach, magnetic microbead

core elements can be detected by a GMR or TMR device positioned above or in close proximity. The presence and number of microbeads are reflected as a measurable change in sensor resistance, enabling quantitative bead counting. Because this method relies on magnetic signals rather than electrical charges, it is not affected by Debye screening in physiological solutions and offers very high sensitivity. As a result, GMR/TMR-based detection is particularly sensitive for quantitative bead counting and bioassays. In general, magnetic detection requires integration of specialized magnetic sensors and involves more complex device fabrication; however, it holds strong potential for advanced biosensing applications.

GMR-based biosensors utilize changes in electrical resistance from variations in local magnetic fields to enable highly sensitive detection of biomolecular interactions. Sun *et al.* have reported the development of a GMR-based biosensing system specifically designed for the separable detection of E. coli O157:H7, a pathogenic strain of Escherichia coli commonly associated with foodborne illness.[56] This work, as in Figure 11a and 11b, demonstrates how magnetic labeling and GMR sensing can be combined to identify whole bacteria with improved sensitivity and specificity.

Wang *et al.* have focused on how GMR biosensors perform when combined with superparamagnetic nanoparticles (MNPs) used as magnetic labels, as in Figure 11c.[57] They built a GMR detection system with a 64-sensor array fabricated by photolithography, allowing multiplexed measurements. DNA hybridization was used as a model assay: probe DNA was immobilized on the sensor surface, and target DNA hybridization brought different types of MNPs close to the surface for GMR measurement.

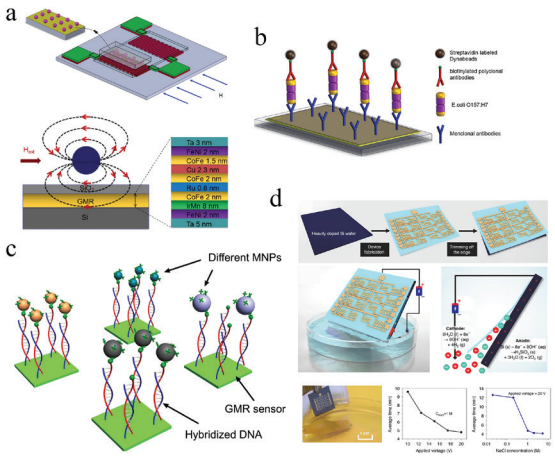


Figure 11: GMR magnetic detections (a) and (b) Microfabricated GMR sensor combined with double antibody sandwich immunoassay and streptavidin-biotin binding assay. Reprinted with permission from Biosens. Bioelectron.[56] Copyright 2016 Elsevier. (c) Magnetoresistive performance and comparison of supermagnetic nanoparticles on a GMR sensor-based detection system. Reprinted with permission from Sci. Rep[57] (d) Typical electrochemical delamination process of a GMR sensor. Reprinted with permission from Microsyst. Nanoeng.[58]

Recent progress in giant magnetoresistance (GMR) sensors has moved the technology beyond rigid chips toward printable, flexible, and skin-like (eSkin) devices. New materials and fabrication methods now allow GMR sensors to be made thinner, lighter, and more flexible, while still maintaining high sensitivity and fast response in real-world applications. Zhang *et al.* have developed an ultrathin, flexible giant magnetoresistive (GMR) sensor array that forms the core of a magnetoelectronic e-skin, as in Figure 11d.[58] An electrochemical delamination process was used, which enables rapid, non-destructive peeling of the ultrathin magnetic sensor array from a rigid substrate onto a flexible platform. These ultrathin, flexible, and skin-like GMR sensors are powerful for next-generation biosensing and wearable diagnostics.

■ Conclusion

The rapid evolution of integrated circuit (IC) fabrication technology has fundamentally promoted the development of biological field-effect transistors (Bio-FETs). Continuous scaling of MOSFET dimensions, progress in EUV lithography, and the widespread adoption of three-dimensional (3D) device architectures such as 3D NAND have enabled unprecedented integration density and device performance. These developments point toward a natural transition of Bio-FETs from planar designs to 3D sensing architectures, where vertical stacking and increased active surface area can further enhance sensitivity and throughput. As device dimensions approach the length scales of biomolecules, Bio-FETs are no longer limited to ensemble-averaged measurements, making single biomolecular or single-event detection increasingly feasible and establishing Bio-FETs as a powerful platform for next-generation biosensing and early disease diagnostics.

Magnetic microbead-based strategies offer significant advantages for Bio-FET applications. Magnetic microbeads provide convenient bioconjugation and enable selective capture of target biomolecules. Their motion and spatial positions can be precisely controlled using microcoils or microsolenoids. Such magnetic manipulation significantly improves detection efficiency and signal reliability, while remaining compatible with microfluidic and on-chip integration. Developments in the hard disk drive (HDD) industry have shown that GMR sensors can achieve ultrahigh sensitivity, gigahertz-speed operation, and large-scale manufacturability. Translating this maturity to biosensing suggests a future in which a single chip or "one-disk" platform could rapidly test multiple biomarkers in parallel, using magnetic beads as universal labels and GMR or Bio-FET readout for quantitative detection. Such convergence of semiconductor electronics, magnetic sensing, and biofunctional materials points toward scalable and programmable biosensing architectures with transformative potential in precision early disease detection.

■ Acknowledgments

I am deeply grateful for the step-by-step guidance and support of Dr. Hamidreza Shaye, whose mentorship in scientific writing and thoughtful suggestions were essential to the completion of this paper.

DOI: 10.36838/IJHSR819.10

References

1. International Roadmap for Devices and Systems (IRDS). *IRDS 2023 Edition*; IEEE, 2023. https://irds.ieee.org.
2. Alberts, B.; Johnson, A.; Lewis, J.; Morgan, D.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 7th ed.; Garland Science, 2022.
3. Lodish, H.; Berk, A.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Bretscher, A.; Ploegh, H.; Matsudaira, P. *Molecular Cell Biology*, 9th ed.; W. H. Freeman, 2021.
4. Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry: Life at the Molecular Level*, 5th ed.; Wiley, 2019.
5. Janeway, C. A.; Travers, P.; Walport, M.; Shlomchik, M. *Janeway's Immunobiology*, 9th ed.; Garland Science, 2021.
6. Flint, S. J.; Enquist, L. W.; Racaniello, V. R.; Skalka, A. M. *Principles of Virology*, 5th ed.; ASM Press, 2020.
7. Bergveld, P. Development of an Ion-Sensitive Solid-State Device for Neurophysiological Measurements. *IEEE Trans. Biomed. Eng.* **1970**, *17* (1), 70–71. https://doi.org/10.1109/TBME.1970.4502688.
8. Poghossian, A.; Schöning, M. J. Label-Free Sensing of Biomolecules with Field-Effect Devices for Clinical Applications. *Biosens. Bioelectron.* **2009**, *25* (1), 1–14. https://doi.org/10.1016/j.bios.2009.02.028.
9. Stern, E.; Wagner, R.; Sigworth, F. J.; Breaker, R.; Fahmy, T. M.; Reed, M. A. Label-Free Immunodetection with CMOS-Compatible Semiconducting Nanowires. *Nano Lett.* **2007**, *7* (11), 3405–3409. https://doi.org/10.1021/nl071400y.
10. Schöning, M. J.; Poghossian, A. Recent Advances in the Development of Biosensors Based on Field-Effect Transistors. *Nat. Nanotechnol.* **2008**, *3* (12), 689–694. https://doi.org/10.1038/nnano.2008.282.
11. Ravariu, C.; Manea, E.; Pârvulescu, C.; Dima, G. Nanoporous Layer Integration for the Fabrication of ISFET and Related Transistor-Based Biosensors. *Chemosensors* **2025**, *13* (8), 316. https://doi.org/10.3390/chemosensors13080316.
12. Yang, H.; Sakata, T. Molecular-Charge-Contact-Based Ion-Sensitive Field-Effect Transistor Sensor in Microfluidic System for Protein Sensing. *Sensors* **2019**, *19* (15), 3393. https://doi.org/10.3390/s19153393.
13. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R. L.; Soerjomataram, I.; Jemal, A. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **2024**, *74* (3), 229–263. https://doi.org/10.3322/caac.21834.
14. Kumar, S.; Senapati, S.; Chang, H.-C. Extracellular Vesicle and Lipoprotein Diagnostics (ExoLP-Dx) with Membrane Sensor: A Robust Microfluidic Platform to Overcome Heterogeneity. *Biomicrofluidics* **2024**, *18* (4), 041301. https://doi.org/10.1063/5.0218986.
15. Kesler, V.; Murmann, B.; Soh, H. T. Going beyond the Debye Length: Overcoming Charge Screening Limitations in Next-Generation Bioelectronic Sensors. *ACS Nano* **2020**, *14* (12), 16194–16201. https://doi.org/10.1021/acsnano.0c08622.
16. Jiawei, S.; Zhi, C.; Kewei, T.; Xiaoping, L. Magnetic Bead-Based Adsorption Strategy for Exosome Isolation. Front. Bioeng. *Biotechnol.* **2022**, *10*, 942077. https://doi.org/10.3389/fbioe.2022.942077.
17. Kozhevnikova, D.; Chernyshev, V.; Yashchenok, A. Progress in Isolation and Molecular Profiling of Small Extracellular Vesicles via Bead-Assisted Platforms. *Biosensors* **2023**, *13* (7), 688. https://doi.org/10.3390/bios13070688.
18. Solovicová, V.; Ďatková, A.; Bertók, T.; Kasák, P.; Vikartovská, A.; Lorencová, L.; Tkac, J. Advances in Magnetic Affinity-Based Isolation/Detection of Exosomes for Robust Diagnostics. *Microchim. Acta* **2025**, *192* (4), 206. https://doi.org/10.1007/s00604-025-07048-6.
19. Wu, C. C.; Tsantilas, K. A.; Park, J.; Plubell, D.; Sanders, J. A.; Naicker, P.; Govender, I.; Buthelezi, S.; Stoychev, S.; Jordaan, J.; Merrihew, G.; Huang, E.; Parker, E. D.; Riffle, M.; Hoofnagle, A. N.; Noble, W. S.; Poston, K. L.; Montine, T. J.; MacCoss, M. J. Enrichment of Extracellular Vesicles Using Mag-Net for the Analysis of the Plasma Proteome. *Nat. Commun.* **2025**, *16* (1), 5447. https://doi.org/10.1038/s41467-025-60595-7.
20. Cao, S.; Sun, P.; Xiao, G.; Tang, Q.; Sun, X.; Zhao, H.; Zhao, S.; Lu, H.; Yue, Z. ISFET-based Sensors for (Bio)Chemical Applications: A Review. *Electrochem. Sci. Adv.* **2023**, *3* (4), e2100207. https://doi.org/10.1002/elsa.202100207.
21. Sakata, T. Signal Transduction Interfaces for Field-Effect Transistor-Based Biosensors. *Commun. Chem.* **2024**, *7* (1), 35. https://doi.org/10.1038/s42004-024-01121-6.
22. Pandey, M.; Bhaiyya, M.; Rewatkar, P.; Zalke, J. B.; Narkhede, N. P.; Haick, H. Advanced Materials for Biological Field-Effect Transistors (Bio-FETs) in Precision Healthcare and Biosensing. *Adv. Healthc. Mater.* **2025**, *14* (13), 2500400. https://doi.org/10.1002/adhm.202500400.
23. Rai, H.; Singh, K. R.; Natarajan, A.; Pandey, S. S. Advances in Field-Effect Transistor-Based Electronic Devices Integrated with CMOS Technology for Biosensing. *Talanta Open* **2025**, *11*, 100394. https://doi.org/10.1016/j.talo.2024.100394.
24. Sadighbayan, D.; Hasanzadeh, M.; Ghafar-Zadeh, E. Biosensing Based on Field-Effect Transistors (FET): Recent Progress and Challenges. *TrAC Trends Anal. Chem.* **2020**, *133*, 116067. https://doi.org/10.1016/j.trac.2020.116067.
25. Lee, J.; Jang, J.; Choi, B.; Yoon, J.; Kim, J.-Y.; Choi, Y.-K.; Myong Kim, D.; Hwan Kim, D.; Choi, S.-J. A Highly Responsive Silicon Nanowire/Amplifier MOSFET Hybrid Biosensor. *Sci. Rep.* **2015**, *5* (1), 12286. https://doi.org/10.1038/srep12286.
26. Manimekala, T.; Sivasubramanian, R.; Dharmalingam, G. Nanomaterial-Based Biosensors Using Field-Effect Transistors: A Review. *J. Electron. Mater.* **2022**, *51* (5), 1950–1973. https://doi.org/10.1007/s11664-022-09492-z.
27. Panahi, A.; Sadighian, D.; Forouhi, S.; Ghafar-Zadeh, E. Recent Advances of Field-Effect Transistor Technology for Infectious Diseases. *Biosensors* **2021**, *11* (4), 103. https://doi.org/10.3390/bios11040103.
28. Liu, K.-H.; Lin, H.-Y.; Thomas, J. L.; Chen, C.-Y.; Chen, Y.-T.; Chen, C.-Y.; Yang, C.-H.; Lee, M.-H. Sensing of C-Reactive Protein Using an Extended-Gate Field-Effect Transistor with a Tungsten Disulfide-Doped Peptide-Imprinted Conductive Polymer Coating. *Biosensors* **2022**, *12* (1), 31. https://doi.org/10.3390/bios12010031.
29. Janićijević, Ž.; Nguyen-Le, T.-A.; Baraban, L. Extended-Gate Field-Effect Transistor Chemo- and Biosensors: State of the Art and Perspectives. *Nanotechnol.* **2023**, *3–4*, 100025. https://doi.org/10.1016/j.nxnano.2023.100025.
30. Deng, Y.; Liu, L.; Li, J.; Gao, L. Sensors Based on the Carbon Nanotube Field-Effect Transistors for Chemical and Biological Analyses. *Biosensors* **2022**, *12* (10), 776. https://doi.org/10.3390/bios12100776.
31. Li, J.; Wu, D.; Yu, Y.; Li, T.; Li, K.; Xiao, M.-M.; Li, Y.; Zhang, Z.-Y.; Zhang, G.-J. Rapid and Unamplified Identification of COVID-19 with Morpholino-Modified Graphene Field-Effect Transistor Nanosensor. *Biosens. Bioelectron.* **2021**, *183*, 113206. https://doi.org/10.1016/j.bios.2021.113206.
32. McDonnell, S. J.; Wallace, R. M. Atomically-Thin Layered Films for Device Applications Based upon 2D TMDC Materials. *Thin Solid Films* **2016**, *616*, 482–501. https://doi.org/10.1016/j.tsf.2016.08.068.

33. Gao, A.; Lu, N.; Wang, Y.; Li, T. Robust Ultrasensitive Tunneling-FET Biosensor for Point-of-Care Diagnostics. *Sci. Rep.* **2016**, *6* (1), 22554. https://doi.org/10.1038/srep22554.
34. Smaani, B.; Nafa, F.; Benlatrech, M. S.; Mahdi, I.; Akroum, H.; Walid Azizi, M.; Harrar, K.; Kanungo, S. Recent Progress on Field-Effect Transistor-Based Biosensors: Device Perspective. *Beilstein J. Nanotechnol.* **2024**, *15*, 977–994. https://doi.org/10.3762/bjnano.15.80.
35. Pahari, P.; AbdelAll, N.; Mohapatra, S. K.; Das, J. K.; Khouqeer, G. A. Impact of Structural and Material Parameters on Sensitivity of Engineered N-Pocket DGTFET Biosensors. *Front. Phys*. **2026**, *13*, 1757118. https://doi.org/10.3389/fphy.2025.1757118.
36. Chang, H.-C.; Yeo, L. Y. *Electrokinetically Driven Microfluidics and Nanofluidics*; Cambridge Univ. Press: Cambridge, 2010.
37. Murayama, K.; Furukawa, C.; Yamasaki, S.; Ohno, Y.; Kajisa, T.; Nagase, M. Biosensing beyond Debye Screening Length Using Epitaxial Graphene Field-Effect Transistors on SiC Substrate. *Surf. Interfaces* **2024**, *54*, 105279. https://doi.org/10.1016/j.surfin.2024.105279.
38. Vacic, A.; Criscione, J. M.; Rajan, N. K.; Stern, E.; Fahmy, T. M.; Reed, M. A. Determination of Molecular Configuration by Debye Length Modulation. *J. Am. Chem. Soc.* **2011**, *133* (35), 13886–13889. https://doi.org/10.1021/ja205684a.
39. Zhang, A.; Zheng, G. Semiconductor Nanowires for Biosensors. In *Semiconductor Nanowires*; Elsevier, 2015; pp 471–490. https://doi.org/10.1016/B978-1-78242-253-2.00017-7.
40. Cheng, H.-P.; Yang, T.-H.; Wang, J.-C.; Chuang, H.-S. Recent Trends and Innovations in Bead-Based Biosensors for Cancer Detection. *Sensors* **2024**, *24* (9), 2904. https://doi.org/10.3390/s24092904.
41. Modh, H.; Scheper, T.; Walter, J.-G. Aptamer-Modified Magnetic Beads in Biosensing. *Sensors* **2018**, *18* (4), 1041. https://doi.org/10.3390/s18041041.
42. Yue, Q.; Wang, Y.; Chen, S.; Li, J. Magnetically Induced Self-Assembly of Magnetic Nanocomposites for Label-Free Electrochemical Detection of CA125. *Biosens. Bioelectron*. **2017**, *89*, 592–598. https://doi.org/10.1016/j.bios.2016.09.041.
43. Fang, X.; Chen, C.; Liu, B.; Ma, Z.; Hu, F.; Li, H.; Gu, H.; Xu, H. A Magnetic Bead-Mediated Selective Adsorption Strategy for Extracellular Vesicle Separation and Purification. *Acta Biomater.* **2021**, *124*, 336–347. https://doi.org/10.1016/j.actbio.2021.02.004.
44. Lin, Y.-H.; Peng, P.-Y. Semiconductor Sensor Embedded Microfluidic Chip for Protein Biomarker Detection Using a Bead-Based Immunoassay Combined with Deoxyribonucleic Acid Strand Labeling. *Anal. Chim. Acta* **2015**, *869*, 34–42. https://doi.org/10.1016/j.aca.2015.03.002.
45. Silverio, V.; Amaral, M.; Gaspar, J.; Cardoso, S.; Freitas, P. Manipulation of Magnetic Beads with Thin Film Microelectromagnet Traps. *Micromachines* **2019**, *10* (9), 607. https://doi.org/10.3390/mi10090607.
46. Pawinanto, R. E.; Yunas, J.; Said, M. Mohd.; Noor, M. Mohd.; Majlis, B. Y. Design and Fabrication of PCB-Based Planar Micro-Coil for Magnetic MEMS Actuator; IEEE.
47. Heuberger, A.; Löchel, B. Optical DUV-Lithography for High Microstructures. *Microsyst. Technol*. **1996**, *3* (1), 1–6. https://doi.org/10.1007/s005420050045.
48. Le, H. T.; Mizushima, I.; Nour, Y.; Tang, P. T.; Knott, A.; Ouyang, Z.; Jensen, F.; Han, A. Fabrication of 3D Air-Core MEMS Inductors for Very-High-Frequency Power Conversions. *Microsyst. Nanoeng.* **2018**, *4* (1), 17082. https://doi.org/10.1038/micronano.2017.82.
49. Le, H. T.; Haque, R. I.; Ouyang, Z.; Lee, S. W.; Fried, S. I.; Zhao, D.; Qiu, M.; Han, A. MEMS Inductor Fabrication and Emerging Applications in Power Electronics and Neurotechnologies. *Microsyst. Nanoeng*. **2021**, *7* (1), 59. https://doi.org/10.1038/s41378-021-00275-w.
50. Lefebvre, O.; Cao, H. H.; Cortés Francisco, M.; Woytasik, M.; Dufour-Gergam, E.; Ammar, M.; Martincic, E. Reusable Embedded Microcoils for Magnetic Nano-Beads Trapping in Microfluidics: Magnetic Simulation and Experiments. *Micromachines* **2020**, *11* (3), 257. https://doi.org/10.3390/mi11030257.
51. Cai, G.; Yang, Z.; Chen, Y.-C.; Huang, Y.; Liang, L.; Feng, S.; Zhao, J. Magnetic Bead Manipulation in Microfluidic Chips for Biological Application. *Cyborg Bionic Syst.* **2023**, *4*, 0023. https://doi.org/10.34133/cbsystems.0023.
52. Yue, Y.; Chen, X.; Wang, J.; Ma, M.; He, A.; Liu, R. Label-Free Electrochemical Biosensor with Magnetically Induced Self-Assembly for the Detection of Cancer Antigen 125. Arab. *J. Chem.* **2023**, *16* (9), 105070. https://doi.org/10.1016/j.arabjc.2023.105070.
53. Yu, Y.; Liang, C.; Wan, Q.-Q.; Jin, D.; Liu, X.; Zhang, Z.; Sun, Z.-Y.; Zhang, G.-J. Integrated FET Sensing Microsystem for Specific Detection of Pancreatic Cancer Exosomal miRNA10b. *Anal. Chim. Acta* **2023**, *1284*, 341995. https://doi.org/10.1016/j.aca.2023.341995.
54. Sun, J.; Xie, X.; Xie, K.; Xu, S.; Jiang, S.; Ren, J.; Zhao, Y.; Xu, H.; Wang, J.; Yue, W. Magnetic Graphene Field-Effect Transistor Biosensor for Single-Strand DNA Detection. *Nanoscale Res. Lett.* **2019**, *14* (1), 248. https://doi.org/10.1186/s11671-019-3048-1.
55. Zhu, X.; Cheng, K.; Ding, Y.; Liu, H.; Xie, S.; Cao, Y.; Yue, W. Magnetically Controlled Graphene Field-Effect Transistor Biosensor for Highly Sensitive Detection of Cardiac Troponin I. *Discov. Nano* **2023**, *18* (1), 106. https://doi.org/10.1186/s11671-023-03886-6.
56. Sun, X.; Lei, C.; Guo, L.; Zhou, Y. Separable Detecting of Escherichia Coli O157H:H7 by a Giant Magneto-Resistance-Based Bio-Sensing System. *Sens. Actuators B Chem*. **2016**, *234*, 485–492. https://doi.org/10.1016/j.snb.2016.04.183.
57. Wang, W.; Wang, Y.; Tu, L.; Feng, Y.; Klein, T.; Wang, J.-P. Magnetoresistive Performance and Comparison of Supermagnetic Nanoparticles on Giant Magnetoresistive Sensor-Based Detection System. *Sci. Rep*. **2014**, *4* (1), 5716. https://doi.org/10.1038/srep05716.
58. Zhang, J.; Jin, Z.; Chen, G.; Chen, J. An Ultrathin, Rapidly Fabricated, Flexible Giant Magnetoresistive Electronic Skin. *Microsyst. Nanoeng.* **2024**, *10* (1), 109. https://doi.org/10.1038/s41378-024-00716-2.

Author

Melody Wang is a rising senior high school student. She has a strong passion for science and engineering, especially in the fields of biology, biomedical technology, and innovative healthcare solutions. She is deeply interested in exploring how novel engineering principles can be applied to solve real-world medical challenges and is motivated to contribute in meaningful ways through continued learning, research, and collaboration.

DOI: 10.36838/IJHSR819.10

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

The Path to Green AI – Challenges and Solutions

Nikhil Bhadra

Staples High School, 70 North Avenue, Westport, CT-06880, USA; nbhadra60@gmail.com

ABSTRACT: Recent growth in Artificial Intelligence (AI) systems has caused an unprecedented demand for large-scale, high-speed computing resources. This has led to increased concerns about the sustainability of AI development and its environmental impact. This study examines different strategies through a literature review and uses key metrics such as efficiency, cost, and feasibility to compare them. The findings revealed that software solutions such as algorithm optimization, the use of small language models, and data curation achieve the highest efficiency by reducing computational demand. Hardware solutions provide a baseline level of efficiency while workload optimization minimizes energy consumption through distributed learning and energy-aware scheduling. Finally, infrastructure solutions offer a downstream option to minimize energy use in data centers through efficient thermal management. The study also emphasizes challenges in standardizing carbon reporting and establishing enforceable regulatory systems and stresses that AI researchers, data center engineers, and energy policy experts must collaborate to translate technical efficiency gains into real-world emissions reductions for sustainable AI development.

KEYWORDS: Green AI, Hardware Optimization, Algorithm Optimization, Small Language Model, Data Curation, Workload Optimization, Grid Optimization.

Introduction

AI has become an important component of modern society. As AI capabilities evolve, the power required to train large models has increased faster than improvements in energy efficiency, causing a significant impact on the environment.[1]

This paper addresses the critical question: Can we keep up with the growing demand for AI without negatively impacting the environment?

Green AI has emerged as the field to address this challenge. It refers to the development and use of AI systems that reduce energy consumption, carbon emissions, and computational waste while maintaining strong performance. It aims to address the rising environmental and economic costs of training and deploying large AI models. Although Green AI is still an emerging field, it has gained increasing attention from researchers, technology companies, and policymakers.

The present study aims to investigate various green AI strategies by comparing them against key metrics such as efficiency, cost, and feasibility. By synthesizing current strategies and reviewing their merits and shortcomings, this work will contribute to current efforts in Green AI. While previous literature has extensively discussed each strategy in isolation, this study aims to provide a comparative analysis of the strategies through both literature review and qualitative diagnostics, to examine their combined benefits in reducing energy consumption.

By analyzing these approaches alongside macro- and policy-level discussions, this study aims to emphasize the distinct natures of Green AI strategies and their collective contribution to sustainable AI development.

Green AI Solutions

Sustainable AI development is a multi-solution approach built on different concepts and technologies. They can be categorized into three main areas: technical, operations, and infrastructure. We will look at the different options under each category and compare them.

Technical Solutions:

Technical solutions include a combination of hardware and software solutions that can provide significant energy savings.

1. Hardware Optimization:

Hardware optimization covers efficiency techniques such as specialized accelerators, low-power AI chips, and application-specific integrated circuits (ASICs).

Specialized Hardware Accelerators:

Special hardware accelerators, like Tensor Processing Units (TPUs), achieve much higher performance per watt than regular GPUs by optimizing the matrix-multiplication architecture. Experimental and production deployments have shown that TPUs can improve inference performance per watt by up to 15×, with gains of 25–30× TOPS/Watt across different neural network architectures.[2]

Low-Power AI Chips and Edge Accelerators:

Low-power AI hardware such as NVIDIA's Jetson Nano and Google's Coral Edge TPU can do about 4 trillion operations per second (TOPS) while using only 2 watts of power.[3] Compared to regular 64-bit ARM CPUs, Edge TPUs consume less energy while keeping good classification accuracy for deep feedforward neural networks.

Application-Specific Integrated Circuits (ASICs):

ASICs are custom-made for specific AI tasks, enabling significant improvements in both performance and power con-

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.11

sumption. ASICs use much less energy than GPUs for steady, repetitive AI tasks.[4]

Table 1 provides a comparison of different hardware optimization strategies. Hardware optimization by itself only provides baseline efficiency. We will discuss next the different software solutions that reduce the need for general-purpose computation and make hardware solutions more effective.

Table 1: Comparison of different hardware optimization strategies.

Hardware Type	Primary Use Case	Energy Efficiency	Flexibility	Cost
GPU	Training & inference	Low	High	Medium
TPU / Accelerator	AI-specific workloads	High	Medium	Medium
Edge AI Chips	On-device inference	Very High	Low	Low
ASIC	Fixed AI workloads	Very High	Very Low	Very High

Table 1 - Trade-off between energy efficiency, flexibility, and cost for different hardware optimization solutions. ASICs and edge accelerators deliver the highest energy efficiency. However, they offer limited flexibility and are mainly used for well-defined workloads. GPUs, on the other hand, offer high flexibility but incur significant energy costs.

2. Algorithm Optimization:

As AI systems continue to scale, algorithm-level optimization, for both training and inference, has emerged as a critical pathway toward sustainable AI.

Training Optimization:

Energy efficiency during the training phase can be improved through data-level, model-level, and algorithm-level optimization strategies.[5] At the data level, **Stochastic mini-batch dropping (SMD)** reduces computational overhead by selectively skipping less informative mini-batches during training. At the model level, **Input-dependent selective layer update (SLU)** minimizes redundant computation by updating only the most relevant layers for a given input. At the algorithm level, **Predictive sign gradient descent (PSG)** reduces training energy costs by approximating gradient directions rather than computing exact gradient values.[5] Complementary techniques, such as **Adaptive learning rate scheduling** and **Early stopping**, further reduce energy usage by avoiding unnecessary computation.[5]

Inference Optimization:

Inference optimization strategies include **Parameter pruning**, which removes redundant or low-importance weights from neural networks. **Quantization** lowers numerical precision, enabling faster execution with minimal impact on accuracy. Additional techniques, such as **Convolutional filter compression** and **Matrix factorization**, simplify network operations, while **Network architecture search (NAS)** automatically identifies model structures optimized for efficiency.[6] **Knowledge distillation** further improves efficiency by transferring knowledge from large "teacher" models to smaller "student" models, enabling comparable performance at a fraction of the computational and energy cost.[6]

Table 2 provides a comparison of the different algorithm optimization solutions. While algorithm optimization is beneficial, it is still extensive in the context of large language models. In the next section, we will discuss small language models for specific tasks, which may significantly limit the scope of training by restricting the model's size.

Table 2: Comparison of different training and inference strategies.

Phase	Optimization Technique	Primary Objective	Energy Efficiency
Training	Stochastic Mini-Batch Dropping (SMD)	Reduce the number of training samples per iteration	Medium
Training	Selective Layer Update (SLU)	Avoid unnecessary layer updates	High
Training	Predictive Sign Gradient Descent (PSG)	Simplify gradient computation	Medium
Inference	Parameter Pruning	Remove redundant weights	High
Inference	Quantization	Reduce numerical precision	High
Inference	Knowledge Distillation	Transfer knowledge to smaller	High

Table 2 - Primary objective and energy efficiency of different training and inference optimization solutions. Most strategies provide significant energy savings when properly implemented and when the primary objective is clear.

3. Small Language Models:

Small language models reduce computing power, memory use, and training time by reducing the number of parameters and optimizing architectural design. DistilBERT, a distilled version of BERT, is 40% smaller and retains 97% of its language understanding while being 60% faster.7 SLMs are increasingly being used in on-device intelligence systems, requiring less data transfer and thereby using less energy and providing better data protection.

Training Efficient Models from Scratch:

One approach to obtaining SLMs involves training models from scratch using efficiency-oriented architectures. There are three strategies: **1. Efficient feed-forward networks** (FFNs), which reduce parameter redundancy in dense layers; **2. Efficient attention mechanisms**, which mitigate the quadratic complexity of standard self-attention through sparse or linear approximations; and **3.Attention-free architectures**, which replace attention mechanisms with alternative sequence modeling techniques.[8]

Model Compression from Large Language Models:

An alternative approach derives SLMs from pretrained LLMs using model compression techniques such as **Quantization**, **Pruning**, and **Knowledge Distillation**, as discussed in the Algorithm Optimization section.[8]

Collaborative SLM–LLM Frameworks:

SLMs and LLMs can be combined, with LLMs providing general thinking and language knowledge, while SLMs act as experts for domain-specific tasks.[8] This division reduces the need for repeated fine-tuning of large models and lowers overall computational cost.

Optimizing training and limiting model size can save significant computing power. In the next section, we will review another strategy for curating training data and reducing its size to improve training efficiency.

4. Data Curation:

Large Language Models (LLMs) are commonly trained on massive web-scale corpora such as C4, Refined Web, and The Pile. However, prioritizing data quality through curated, smaller datasets can enable effective LLM training with lower energy consumption.[9]

Data Curation as an Energy Saving Strategy:

Data curation focuses on extracting high-quality and informative subsets from large corpora. By doing so, models can be trained on smaller yet more representative datasets that facilitate learning important linguistic patterns without unnecessary tokens. Subramanyam *et al.*[10] through experimental work, proposed a new quality-aware scaling law that extends the widely used Chinchilla framework, predicting pretraining loss as a joint function of model size, data volume, and data quality.

Table 3 compares the size and the number of hours needed to train a model. Approximately 15 to 20 domain-specific models can be trained with the same amount of energy needed to train a single LLM. Focusing on building smaller, domain-specific models rather than general-purpose LLMs can result in significant energy savings.

Table 3: Comparison of hours and energy needed to train models of different sizes.

Dataset Type	Size (B Tokens)	GPU Hours	Estimated Energy (MWh)	Model Type
Large-scale Web Dataset (Uncurated)	300	50,000	2,500	Full C4/Pile-style corpora
Curated High-Quality Subset	50	8,000	400	SLM-evaluated, smaller but informative
Task-Specific Curated Dataset	20	3,000	150	Domain-specific, highly filtered

Table 3 - Estimated GPU hours and associated energy usage for training different types of datasets. Large-scale web datasets are typically used to train LLMs like GPT-3 (300B tokens). SLMs and other domain-specific models need much smaller datasets and considerably less energy for training.

Operations Solutions:

In this section, we will discuss the benefits of distributed learning and proactive scheduling to maximize the efficiency of training models.

1. Workload Optimization:

Federated Learning:

Federated Learning (FL) creates machine learning models from data distributed across many devices. Each device uses its local data to update the model, and only that update is sent back to the server.[11] FL also makes it simpler to follow data privacy and security rules as data stays localized.[12]

Energy-Aware Scheduling:

Energy-aware scheduling techniques consider power consumption, cooling requirements, and workload characteristics to optimize energy use.[13] They maintain workload profiles to schedule tasks during periods when energy demand is low or when large amounts of renewable energy are available. P. Krishnaram did a study where, at a peak utilization setting of 0.69%, the proposed workload optimization framework achieved a 13% time-saving efficiency compared to static scheduling, corresponding to an estimated energy savings of 364 MW for LLMs with over 10 billion parameters.[14]

Infrastructure Solutions:

Infrastructure solutions provide the most downstream optimization solution in the Green AI framework. In the following sections, we will discuss strategies that can provide efficient solutions for managing the increasing workload in AI data centers.

1. Efficient Heat Management:

Cooling systems at data centers consume nearly 50% of electricity. Techniques such as immersion and liquid cooling are increasingly being adopted in place of traditional air-cooling systems to manage the high cooling requirements and associated power demand in AI data centers.

Cooling Strategies:

Immersion cooling submerges servers or components in a thermally conductive, electrically insulating dielectric fluid. Heat is directly absorbed by the fluid, eliminating the need for fans or air-based heat removal. Immersion cooling can offer energy efficiency gains of 40% compared to conventional air cooling.[15]

Liquid cooling circulates a coolant, often a mixture of ethylene glycol and deionized water, directly to heat-generating components via cold plates. Heat generated by the equipment is efficiently conducted through the liquid, which is then transported to the heat exchanger, where it is dissipated via the external cooling system. Liquid cooling can offer energy efficiency gains of 60-90% compared to conventional air cooling.[16]

Other Heat Management Strategies:

Microfluidic cooling delivers coolant directly inside silicon chips via etched microchannels. By removing heat at its source, it dramatically reduces thermal resistance compared to conventional cold-plate solutions. Rajan *et al.*[17] used an Intel Core i7-8700K CPU with micropin fins etched directly on the back and overclocked it to dissipate up to 215W of power while cooled by room-temperature de-ionized (DI) water. The experiment demonstrated up to 44.4% reduction in junction-to-inlet thermal resistance while using only 0.3× of volumetric coolant flow per Watt of power dissipated in the CPU compared to a conventional cold plate.[17]

Waste Heat Reuse (WHR), on the other hand, improves data center sustainability by redirecting excess heat to external applications, such as district heating, hot water supply, or industrial processes.[18] Case studies like Stockholm Data Parks (Sweden) and the University of Notre Dame (USA) demonstrate that WHR not only reduces environmental impact but also creates economic value from otherwise wasted thermal energy.

The largest gains in cooling efficiency come from more advanced techniques, such as liquid and immersion cooling, as shown in Table 4. However, they provide diminishing re-

DOI: 10.36838/IJHSR819.11

turns when considered in isolation, as these gains come with increased system complexity and deployment challenges. Cooling in general is a downstream optimization layer whose effectiveness depends on upstream computational demand.

Table 4: Comparison of different data center cooling strategies.

Cooling Method	Efficiency	Complexity	Sustainability	Typical Use Case
Air Cooling	Low	Low	Minimal	Standard servers
Liquid Cooling	Very High (approx. 60% energy savings)	Medium	Moderate	AI racks
Immersion Cooling	High (20–40% energy savings)	Medium	Moderate	AI clusters
Microfluidic Cooling	Very High	High	Moderate	Next-gen AI chips

Table 4- Complexity and Sustainability of various data center cooling strategies. Liquid Cooling and Immersion Cooling provide the best balance of complexity and sustainability, whereas next-generation solutions like Microfluidic Cooling are still technically complex and difficult to implement at scale.

2. Alternate Strategies:

As rack power densities in AI data centers continue to increase, alternative approaches that completely reimagine data center placement and cooling environments are gaining attention. **1. Underwater data centers** leverage the naturally low and stable temperatures of deep ocean environments to dissipate their heat efficiently.[19] **2. Space-based data centers** deploy computational infrastructure into Earth's orbit using solar power as a primary energy source and radiative cooling to dissipate heat directly into space.[20] **3. High-altitude platforms** benefit from passive thermal dissipation due to low ambient temperatures, as they operate in the stratosphere at altitudes of 18–25 km.[21] **4. Intelligent control mechanisms** reduce cooling energy consumption by using machine learning models to predict thermal behavior.[22]

Table 5 illustrates a clear distinction between theoretical efficiency and deployment feasibility for alternate infrastructure solutions. Solutions that can be integrated within existing infrastructure constraints provide the best path to sustainable AI when compared with the most extreme or novel solutions.

Table 5: Comparison of different alternate infrastructure strategies.

Approach	Cooling Demand	Energy Efficiency	Deployment Feasibility
Conventional Terrestrial DC	High	Low	High
Underwater DC	Very Low	High	Medium
Space-Based DC	Minimal	Very High	Low
High-Altitude DC	Low	Moderate	Medium
AI-Driven Cooling	Moderate	High	High

Table 5 - Energy efficiency and deployment feasibility for various alternate infrastructure solutions. Space-based and underwater data centers offer extremely high cooling efficiency. However, they are unlikely to make a significant contribution to near-term sustainability goals due to their low feasibility and high cost. In contrast, AI-driven cooling and high-altitude platforms offer more balanced trade-offs between efficiency and deployability.

■ Comparative Trade-Off Analysis

Each Green AI strategy presents distinct benefits and limitations, and no single approach provides the best optimization solution across efficiency, cost, and feasibility. Rather than competing, these approaches provide the best results when applied collectively within a Green AI framework.

Table 6 compares all major Green AI strategies across multiple dimensions. To maximize overall system sustainability, a layered optimization framework is needed in which software efficiency reduces demand and enables more effective use of hardware and infrastructure solutions.

Table 6: Comparison of different Green AI strategies.

Solution	Efficiency	Feasibility	Cost	Complexity	Current Adoption	References
Hardware Optimization (GPUs, TPUs, ASICs)	High	High	Medium–High	Medium–High	Growing (production use)	[2], [3], [4]
Algorithm Optimization (Training, Inference)	High	High	Low–Medium	Medium	Moderate–High	[5], [6]
Small Language Models (SLMs)	Very High (task-specific)	High	Low–Medium	Medium	Growing	[7], [8]
Data Curation	Medium–High	High	Low–Medium	Medium	Early Growing	[9], [10]
Workload Optimization	Medium	Medium–High	Medium	High	Early	[11], [12], [13], [14]
Efficient Heat Management	High	Medium–High	High	High	Moderate	[15], [16], [17], [18]
Alternate Strategies	Very High (theoretical)	Low–Medium	Very High	Very High	Experimental	[19], [20], [21], [22]

Table 6 - Green AI strategies compared against efficiency, feasibility, cost, and complexity. Hardware optimization, algorithm optimization, and efficient heat management provide the best near-term benefits with a moderate to high level of adoption. Data curation and workload optimization show early promise and are expected to be increasingly adopted. Alternate strategies provide the best efficiency benefits but are limited by extremely high complexity and cost.

■ Macro Analysis

This section analyzes sustainability challenges at the macro level by comparing current and future energy use for the two global AI powers, the US and China.

Table 7 and Figure 1 show the current and projected data center energy consumption in the U.S. and China. In both countries, increased scale and adoption of AI are projected to outpace efficiency gains. The primary energy source in the US is natural gas, and in China it is coal. The same computational workload, therefore, can yield significantly different carbon emissions in these two countries based on their underlying energy infrastructure. The adoption of renewable energy and grid optimization, therefore, must be integrated with technical efficiency to enable sustainable AI in both countries.

Table 7: Annual electricity consumption of data centers in the US and China. Data sourced from Pew Research Center and Bloomberg NEF-reported capacity estimates.

Metric	United States	China
2024 Base Energy Use	183 TWh (Approx. 4% of U.S. total)	Approx. 20 GW aggregate (national)
2030 Projection	Approx. 426 TWh (approx. +133%)	Approx. 40 GW (approx. +100% from 2024)
Share of Electricity Demand	Potentially 8–12% by late decade (industry estimates)	Approx. 3–6% of the national grid by 2030 (industry & policy forecasts)
Contribution to Global Demand Growth	Largest single national driver	Second largest driver

Table 7 - Energy consumption by U.S. and China data centers. In both countries, the energy usage is projected to more than double by 2030. The share of overall electricity demand is still low, but is expected to increase by 2030 in both countries.

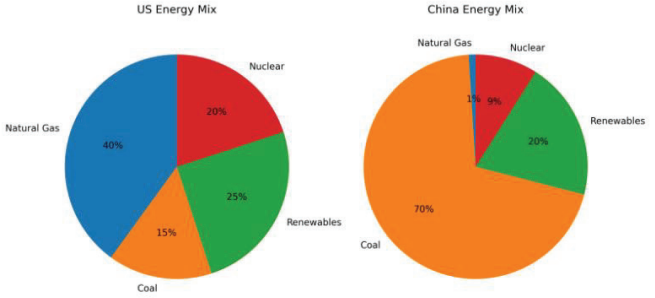


Figure 1: Energy source mix for data centers in the US and China reported by the Pew Research Center and the IEA.

Figure 1 - US data centers get 25% of their energy from renewable sources, while data centers in China get 20%. Natural gas is the largest source of power for US data centers at 40%, while in China, it's coal at 70%. Both countries plan to invest heavily in renewables and nuclear energy for powering AI data centers in the future.

US Strategy:

US energy companies and data center businesses are worried about the strain on the power grid. Delays in upgrading transmission lines are a major obstacle to increasing capacity to support AI data centers. Large companies like Microsoft are promising to pay full energy rates and replace the water used for cooling their data centers. Google and other companies have agreed with energy providers to reduce data center power use during periods of high grid demand. There is growing interest in increasing nuclear and renewable energy use in the US.

China Strategy:

China is focusing on quickly building energy grids and infrastructure to support energy-intensive AI data centers, without incurring long delays due to regulations. Regional plans are increasingly linking data center growth with local renewable energy sources. Areas with abundant renewable energy are given priority for large-scale facilities. Data centers in China are aiming for Power Usage Effectiveness (PUE) benchmarks (e.g., ≤1.25) to improve energy efficiency compared to global standards.

■ Results and Discussion

Implementation of a sustainable AI strategy involves the integration of multiple complementary strategies rather than the deployment of any single solution. Tables 1 to 4 list the different optimization strategies along with their merits and shortcomings. Hardware optimization improves performance per watt and provides baseline efficiency, but is constrained by flexibility and cost. Algorithmic optimization and small language models provide better scalable efficiency gains by directly reducing computational demand. Data curation reduces unnecessary training overhead by prioritizing quality over quantity. Workload optimization extends efficiency into deployment. Cooling and infrastructure strategies act as downstream enablers whose effectiveness depends on the upstream computational workload.

Table 5 highlights that at the macro level, rising global AI energy demand will outpace isolated efficiency gains. Green AI, therefore, must be understood as a unified framework where grid optimization and increased use of renewable energy are as important as software, hardware, and infrastructure optimizations.

Future studies are needed to assess the real-world feasibility and to quantify the combined effects of multiple Green AI strategies deployed simultaneously. Additional research is also needed on standardized reporting of energy and carbon costs. AI researchers, data center engineers, and energy policy experts need to collaborate to translate technical efficiency gains into real-world emissions reductions.

■ Conclusion

The paper is primarily based on a literature synthesis from previous work in this field. Green AI is an evolving field, and sufficient empirical data from experiments is not available to analyze the combined effect of multiple Green AI strategies in real-world settings. Hence, most of the comparative analysis in this paper is based on qualitative diagnostics. The optimization solutions listed mainly focus on technical and infrastructure efficiencies and do not fully quantify lifecycle carbon emissions or economic trade-offs. In addition, rapidly evolving hardware and AI models may render some findings outdated and pose new challenges. Finally, regional comparisons are limited to the U.S. and China, the two major AI powerhouses. Increased AI adoption globally may need a more holistic evaluation of broader energy usage and policy environments.

■ Conclusion

The rapid expansion of AI has placed considerable computational and energy demands, raising major concerns about the sustainability of current AI development practices. Hardware and algorithm optimization, small language models, data curation, workload management, and advanced thermal regulation systems provide distinct but complementary approaches to reducing the environmental impact of AI. The best approach involves implementing a layered optimization framework that incorporates all the solutions discussed, with software efficiency providing the maximum benefit by reducing overall computational demand.

Though progress has been made in the Green AI initiative, major challenges remain in standardizing carbon reporting, ensuring accuracy with efficiency, and establishing enforceable regulatory systems.

Governments, industry, and academia are all investing significant time and effort in Green AI. AI workload will continue to increase, but with the combined effort from all parties, the challenge for sustainable AI development can be met, and future AI growth can be achieved without negatively impacting the environment.

DOI: 10.36838/IJHSR819.11

Acknowledgments

I would like to thank my mentor, Dr. Samuel Lefcourt, PhD, from Johns Hopkins University, for his guidance and support in helping me write my first academic research paper. Dr. Lefcourt provided valuable input and consistently challenged me on my thesis, prompting me to think more deeply, conduct further research, and be more innovative in my approach. I would also like to thank Joe Xiao for guiding me through the editing process. Finally, I want to acknowledge my parents; without their encouragement, I would not have been able to complete this paper.

References

1. Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2019, pp. 3645–3650, doi:10.18653/v1/P19-1355.
2. Pothukuchi, Nikhila. (2025). "Understanding Tensor Processing Units: The Specialized Hardware Revolutionizing AI Computing." International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 11. 2349-2357. 10.32628/CSEIT23112565.
3. Noorbin, Faezeh & Layeghy, Siamak & Kusy, Brano & Jurdak, Raja & Portmann, Marius. (2023). "Exploring Edge TPU for deep feed-forward neural networks." Internet of Things. 22. 100749. 10.1016/j.iot.2023.100749.
4. Wang, Yujie. (2023). "Artificial-Intelligence integrated circuits: Comparison of GPU, FPGA, and ASIC." Applied and Computational Engineering. 4. 99-104. 10.54254/2755-2721/4/20230358.
5. Wang, Yue, Jiang, Ziyu, Chen, Xiaohan *et al.* "E2-Train: Training State-of-the-art CNNs with Over 80% Energy Savings." Advances in Neural Information Processing Systems, vol. 32, 2019, pp. 5139–5151. NeurIPS 2019, arXiv:1910.13349
6. Goel, Abhinav, Tung, Caleb, Lu, Yung-Hsiang *et al.* "A Survey of Methods for Low-Power Deep Learning and Computer Vision." arXiv:2003.11066 (2020).
7. Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. "DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter." Proceedings of the 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing (EMC^2), Association for Computational Linguistics, 2019, pp. 36–40. arXiv:1910.01108, 2020.
8. Zhang, Qin, Ziqi Liu, and Shirui Pan. "The Rise of Small Language Models." Shenzhen University and Griffith University, 2025.
9. Sajith, Aryan, and Krishna Chaitanya Rao Kathala. "Is Training Data Quality or Quantity More Impactful to Small Language Model Performance?" Neural Computing for Advanced Applications, Springer Nature Singapore, 2025, pp. 75–87. Crossref, https://doi.org/10.1007/978-981-95-3739-6_6.
10. Subramanyam, Anirudh, Yuxin Chen, and Robert L. Grossman. "Scaling Laws Revisited: Modeling the Role of Data Quality in Language Model Pretraining." arXiv, Sept. 2025.
11. McMahan, Brendan, Moore, Eider, Ramage, Daniel *et al.* "Communication-Efficient Learning of Deep Networks from Decentralized Data." Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273–1282. arXiv:1602.05629, 2016.
12. Yang, Qiang, Yang Liu, Tianjian Chen, and Yongxin Tong. "Federated Machine Learning: Concept and Applications." ACM Transactions on Intelligent Systems and Technology, vol.10, no.2, 2019, pp.1–19, doi:10.1145/3298981.
13. Dayarathna, Miyuru & Wen, Yonggang & Fan, Rui. (2015). "Data Center Energy Consumption Modeling: A Survey." IEEE Communications Surveys & Tutorials. 1-1. 10.1109/COMST.2015.2481183.
14. P. C. M. Krishnaram, "Green AI: Optimizing Energy Efficiency of Workloads for Sustainable Data Centers." 2025 IEEE Green Technologies Conference (GreenTech), Wichita, KS, USA, 2025, pp. 148-152, doi: 10.1109/GreenTech62170.2025.10977613.
15. Pambudi, Nugroho Agung, Sarifudin, Alfan, Firdaus, Ridho Alfan *et al.* "The Immersion Cooling Technology: Current and Future Development in Energy Saving." Alexandria Engineering Journal, vol. 61, no. 12, 2022, pp. 9509–9527, doi:10.1016/j.aej.2022.02.059
16. Junliang Liu, Xiaoya Zhang, Lique Shi. "Overview of Liquid Cooling Technology Development in Data Centers." Academic Journal of Science and Technology, October 2025 17(2):33-36, doi: 10.54097/p00b6t25
17. S. Kochupurackal Rajan, B. Ramakrishnan, H. Alissa, W. Kim, C. Belady, and M. S. Bakir, "Integrated Silicon Microfluidic Cooling of a High-Power Overclocked CPU for Efficient Thermal Management." in IEEE Access, vol. 10, pp. 59259-59269, 2022, doi: 10.1109/ACCESS.2022.3179387
18. Yuan, Xiaolei, Liang, Yumin, Hu, Xinyi *et al.* "Waste Heat Recoveries in Data Centers: A Review." Renewable and Sustainable Energy Reviews, vol. 188, 2023, article 113777, doi:10.1016/j.rser.2023.113777
19. Periola, A.A., Alonge, A.A. & Ogudo, K.A. "Heat Wave Resilient Systems Architecture for Underwater Data Centers." Sci Rep 12, 17161 (2022). https://doi.org/10.1038/s41598-022-21293-2
20. Periola, A.A., Kolawole, M.O. "Space-Based Data Centers: A Paradigm for Data Processing and Scientific Investigations." Wireless Pers Commun 107, 95–119 (2019)
21. W. Abderrahim, O. Amin, and B. Shihada, "How to Leverage High Altitude Platforms in Green Computing?" IEEE Communications Magazine, vol. 61, no. 7, pp. 134-140, July 2023, doi: 10.1109/MCOM.001.2200732.
22. H. Liu, A. Aljbri, J. Song, J. Jiang, and C. Hua, "Research advances on AI-powered thermal management for data centers." in Tsinghua Science and Technology, vol. 27, no. 2, pp. 303-314, April 2022

Author

Nikhil Bhadra is an aspiring entrepreneur with a strong interest in business and finance. Nikhil plans to study at a top US university to pursue a career in financial services. Nikhil is a swim instructor and founder of WaterSafety4Kids, a growing non-profit that sponsors swim lessons for children from underserved communities.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Diverse Biologic Modalities in Systemic Lupus Erythematosus (SLE) Clinical Trials

Phoebe Hu

Academy for Allied Health Sciences, 1776 Raritan Rd, Scotch Plains, NJ, 07076, USA; phoebehu6@gmail.com
Mentor: Trinny Tat

ABSTRACT: Systemic Lupus Erythematosus (SLE) is a challenging chronic autoimmune disease that affects patients with a wide range of symptoms, including life-threatening complications. It is difficult to treat because of its heterogeneous nature. Despite available treatment options, there is no cure currently. Over the last few decades, tremendous efforts have been made to understand the biologic mechanisms causing SLE and to explore novel and effective treatment options to better control and eventually cure the disease. In recent years, therapeutic advancements in SLE, particularly the use of diverse biologic modalities, have been exciting and offer new hope for patients. In this review, emerging treatment options in clinical trials using biologic modalities such as therapeutic proteins, cell therapies, and mRNA-LNP are discussed. Meanwhile, representative manufacturing processes for each modality are described along with the associated considerations, which are important for maintaining a robust and reliable supply chain to ensure patient access. The aim is to provide a comprehensive overview, from exciting biological hypotheses to eventual patient usage.

KEYWORDS: Disease Treatment and Therapies, Systemic Lupus Erythematosus (SLE), Biologic Modalities, Manufacturing Processes.

Introduction

Systemic lupus erythematosus (SLE) is one of the most difficult-to-treat autoimmune diseases. Due to loss of immune tolerance, SLE patients' immune systems mistakenly attack their own cells and tissues, significantly impacting their health. Individuals with this disease experience a lower quality of life, often suffering from symptoms such as fatigue and pain. Furthermore, it can affect multiple organ systems, presenting a wide range of signs and symptoms that could become life-threatening.[1,2]

Clinical heterogeneity makes it hard to diagnose and manage SLE because patients can experience very different symptoms and disease courses.[3] For example, a defining feature of lupus is the presence of antinuclear antibodies (ANA). However, although more than 97% of cases have a positive ANA test, the specificity of this test is only 20%, making it an unreliable marker for confirming SLE on its own.[4] Despite diagnostic challenges, the 2019 EULAR/ACR classification system is generally considered a valuable tool, which has shown high sensitivity and specificity.[5] Using this evaluation system, patients with 10 or more points and meeting at least one clinical criterion are identified as having SLE. Globally, an estimated 3.4 million people are living with SLE.[6] Statistically, SLE primarily affects females, particularly those of childbearing age.[7]

Considering that SLE has no cure currently, there is a significant unmet medical need. Progress in developing effective treatments for SLE has been slow and challenging. To date, only a limited number of drugs have been specifically approved for the treatment of SLE. Similar to other autoimmune diseases, the standard treatment for SLE relies on immunomodulatory drugs, such as antimalarial therapies. Among these, hydroxychloroquine (HCQ), approved by the FDA in 1955, is currently considered one of the primary drugs for SLE treatment.[8] Next, belimumab was approved in 2011 as the first new lupus drug in over 50 years, which selectively inhibits B-cell activating factor (BAFF), also known as BlyS.[9] SLE patients have an elevated level of this protein in their bodies. Belimumab can neutralize soluble BAFF and then disrupt the survival and maturation signals required for B cells, effectively reducing the population of autoreactive B cells and the subsequent production of harmful autoantibodies. Then, anifrolumab was approved in 2021, which works by binding to the type I interferon receptor (IFNAR1).[10] Most SLE patients tend to overproduce interferons, the signaling proteins that trigger immune system overactivity and persistent inflammation. By blocking this receptor, anifrolumab suppresses the inflammatory signals that lead to organ damage, decreasing the frequency of exacerbations and alleviating symptoms, particularly in the skin and joints. In addition to these three drugs approved for SLE, voclosporin and obinutuzumab were approved for the treatment of lupus nephritis, a major complication of SLE, in 2021 and 2025, respectively.

Despite significant challenges, substantial efforts continue across both academic research and industry to develop new therapies for patients with SLE. In recent years, an upward trend in SLE clinical trial activity has been observed.[11,12] As of February 8, 2026, a search of ClinicalTrials.gov identified approximately 327 active clinical trials related to SLE, including 70 in Phase 2 and 29 in Phase 3. Most recently, one of the leading emerging strategies for treating SLE is known as immune reset or immune reprogramming, which has shown very promising early clinical results.[13]

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.12

With improved understanding of the underlying biological mechanisms causing SLE, scientists can now design diverse therapeutic modalities to slow or stop disease development and progression. As one of the most well-established therapeutic modalities, small-molecule drugs have played an important role in medical treatment since the early 1900s and continue to make significant contributions today.[14] Subsequently, advances in biotechnology, particularly recombinant DNA technology in the 1980s, enabled the development and widespread use of biologic drugs. Among these, monoclonal antibodies have gained particular interest due to their high target specificity and therapeutic versatility. To date, the FDA has approved more than 100 monoclonal antibody therapies for commercial use.[15,16]

Over the last 20 years, new modalities such as cell therapy and gene therapy have emerged and advanced significantly. Cell therapies use living cells, such as immune cells or stem cells, to repair tissues or modify the body's immune system to better fight disease.[17] Gene therapies utilize state-of-the-art technologies to regulate gene activity within cells, offering long-lasting therapeutic effects.[18] More recently, messenger RNA (mRNA) therapies delivered using lipid nanoparticles (LNPs) have gained attention because they enable the body to produce useful proteins without permanently changing DNA.[19]

Despite the urgent need for more effective therapies for SLE, drug development and regulatory approval must go through rigorous processes to ensure patient safety and therapeutic efficacy. To enable clinical evaluation and eventual regulatory approval, the manufacturing strategy is a critical component of the drug development process, ensuring consistent product quality and compliance with regulatory expectations. Notably, the manufacturing processes differ substantially across therapeutic modalities. As therapeutic approaches become increasingly complex, a robust manufacturing process has become essential not only for regulatory approval but also for patient access.

In this review, we focus on several emerging biologic drug candidates currently in clinical trials for the treatment of SLE. They are categorized by their specific modalities (i.e., recombinant protein, cell therapy, and mRNA-LNP) followed by their associated manufacturing processes. By connecting therapeutic innovation with their production requirements, we provide an integrated perspective from scientific hypothesis to supply chain for industrial application.

Discussion

Recombinant Proteins – Clinical Advances:

Due to advancements in recombinant DNA technology, therapeutic proteins have become a major modality for disease treatment, including SLE. In addition to belimumab, anifrolumab, and obinutuzumab mentioned above, a considerable number of protein candidates are being tested in clinical trials, particularly monoclonal antibodies (mAbs) or mAb-like proteins. Below are a few highlights of these candidates in late-stage clinical development.

Ianalumab: Ianalumab is a humanized IgG1 monoclonal antibody currently in phase 3 development (SIRIUS-SLE 1 and SIRIUS-SLE 2), which targets the BAFF receptor (BAFF-R).[20] While belimumab targets the ligand by neutralizing soluble BAFF, ianalumab blocks the BAFF receptor (BAFF-R) expressed on the surface of B cells, thus inhibiting pro-survival and differentiation signals that are important for autoreactive B cells. In addition, ianalumab engages immune effector cells to induce antibody-dependent cellular cytotoxicity (ADCC), resulting in direct depletion of BAFF-R-expressing B cells. Especially, ianalumab is an afucosylated antibody, a protein structural modification designed to enhance ADCC activity.[21] The combination of functional pathway blockade and active B cell depletion can lead to a more profound and sustained reduction of pathogenic B cell populations.

Litifilimab: Litifilimab is a humanized IgG1 monoclonal antibody in Phase 3 clinical trials for SLE (TOPAZ-1 and TOPAZ-2), which targets blood dendritic cell antigen 2 (BDCA2), an inhibitory receptor selectively expressed on plasmacytoid dendritic cells (pDCs).[22] In SLE, immune complexes containing self-derived DNA or RNA can activate endosomal Toll-like receptors, particularly TLR7 and TLR9, in pDCs, resulting in excessive production of type I interferons. When litifilimab binds to BDCA2, it triggers an intracellular inhibitory signaling cascade that suppresses TLR7 and TLR9-mediated pathways.[23] Specifically, it reduces the activation of downstream transcription factors and decreases the production of type I interferons and other pro-inflammatory cytokines.[23] The mechanism of litifilimab is to reduce pDC responses rather than depleting these cells, which can help preserve broader immune function.

Dapirolizumab pegol: Dapirolizumab pegol is a pegylated, Fc-free monoclonal antibody fragment currently in Phase 3 studies (PHOENYCS GO and PHOENYCS FLY), which targets CD40 ligand (CD40L), a key costimulatory pathway involved in immune activation.[24] In SLE, CD40L is overexpressed on activated T cells and promotes pathogenic interactions with CD40 on B cells, dendritic cells, and other antigen-presenting cells, causing autoantibody production and inflammatory cytokine release. By blocking the CD40–CD40L interaction, dapirolizumab pegol inhibits T-cell-dependent B-cell activation and downstream inflammatory signaling.[25] The absence of an Fc region reduces the risk of Fc-mediated platelet activation and associated adverse events observed with earlier anti-CD40L antibodies, enabling safer modulation of this immune pathway in patients with SLE. However, without an Fc domain, the remaining Fab fragments will be quickly cleared. Therefore, a 40 kDa PEG chain is added to increase the hydrodynamic radius.[26] This design slows clearance and extends the half-life while preserving the intended safety profile.[25,27]

Telitacicept: Telitacicept is a recombinant, fully human fusion protein approved in China and is currently being evaluated in a Phase 3 global clinical program for SLE (REMESLE-2). It consists of the extracellular domain of TACI (Transmembrane Activator and CAML Interactor) fused to the Fc portion of human IgG. It acts as a soluble binding receptor

that can simultaneously neutralize two key B-cell survival factors, BAFF and a proliferation-inducing ligand (APRIL).[28] In SLE, elevated BAFF and APRIL levels promote the antibody production of autoreactive B cells. By binding and neutralizing both ligands important for B-cell maturation and plasma cell survival, telitacicept can reduce autoantibody production and immune-mediated inflammation.[29]

In addition to the late-stage drug candidates discussed above, a significant number of innovative therapeutic proteins are currently under early-stage evaluation. We highlight several of these molecules, focusing on their diverse biological pathways and targets as well as unique molecular design.

CM313: CM313 is a humanized IgG1 monoclonal antibody targeting CD38, a surface marker highly expressed on plasma cells and certain activated immune cells.[30] In SLE, long-lived plasma cells play an important role in sustaining autoantibody production and disease activity, which are also relatively resistant to conventional B-cell directed therapies. By binding CD38, CM313 is intended to reduce pathogenic plasma cell populations, which may lead to decreased autoantibody levels and downstream immune complex-mediated inflammation. Unlike anti-CD38 antibodies developed for multiple myeloma (e.g., daratumumab), CM313 is designed for autoimmune applications, where modulation of immune balance is critical. It has been reported that CM313 possesses a significantly higher binding affinity for the CD38 antigen and demonstrates enhanced complement-dependent cytotoxicity (CDC).[30]

Amulirafusp Alfa: Amulirafusp Alfa is a first-in-class bispecific fusion protein that simultaneously targets CD47 and CD20 on B cells.[31] CD20 targeting enables selective binding to B cells, while CD47 blockade disrupts the "don't eat me" signal that normally protects cells from macrophage-mediated phagocytosis. By combining these two mechanisms in a single molecule, Amulirafusp Alfa enhances the clearance of pathogenic B cells while maintaining high target specificity. Therefore, more effective and sustained B-cell depletion, resulting in a deeper immune reset, can be achieved.

A-319: A-319 is a CD19 x CD3 bispecific T cell engager (TCE), which contains an anti-CD3 single-chain variable fragment (scFv) linked to the N-terminal of the heavy chain of an anti-CD19 Fab fragment.[32] By simultaneously binding CD3 on T cells and CD19 on B cells, A-319 brings T cells into proximity to pathogenic B cells, leading to T-cell activation and targeted elimination of CD19-positive cells. However, TCE is also known to potentially trigger cytokine release syndrome (CRS), which could be dangerous. Therefore, a balance must be achieved between the desired deep immune reset and unwanted excessive immune activation. Preliminary results indicate that A-319 is well tolerated, with no cases of high-grade CRS or neurotoxicity observed. The initial data demonstrated rapid B-cell depletion and a reduction in autoantibodies alongside measurable improvements in organ function.[33]

Last but not least, a recent report showed a promising result of the off-label use of teclistamab, a TCE targeting BCMA and CD3 to deplete plasma and B cells, currently approved for relapsed or refractory multiple myeloma.[34] This case involves a 23-year-old woman with refractory SLE unresponsive to multiple conventional therapies.[35] Following a five-week step-up dosing regimen, the patient achieved rapid and complete drug-free remission. This was marked by a SLEDAI-2K score of zero (indicating no active lupus disease) by week six, alongside normal kidney function, the disappearance of anti-dsDNA autoantibodies, and the resolution of all other clinical symptoms.[35] Although mild CRS and hypogammaglobulinemia occurred, both were manageable, and remission remained sustained at the 16-week follow-up.[35]

In summary, a wide array of therapeutic proteins, especially mAbs and mAb-like proteins, are being actively explored as potential treatments for SLE. A significant recent trend is the increased interest in bispecific and multispecific antibody modalities, which are specifically engineered to facilitate more sophisticated immune modulation.[36] Notably, several of these candidates are being repurposed from hematological oncology for application within the autoimmune space.

Recombinant Proteins – Manufacturing Process:

Since recombinant tissue plasminogen activator (rtPA) became the first mammalian cell-derived therapeutic protein approved in 1987, large-scale cell culture has evolved into the primary platform for biopharmaceutical manufacturing.[37] Today, the field has reached a high level of technological maturity, characterized by robust and high-yield production systems along with established regulatory pathways. These platforms now dominate the global biopharmaceutical landscape, including the manufacturing of several blockbuster oncology and immunology drugs.[38]

Cell Line: The first step in preparing for a manufacturing process is to generate a product-specific cell line, which involves engineering a host cell, most commonly a Chinese Hamster Ovary (CHO) cell, to stably and efficiently express the desired therapeutic protein. Over the years, tremendous efforts have been directed toward constructing high-performance host cell lines to maximize productivity and stability.[39] For example, host cells can be engineered to reduce or eliminate their ability to produce the endogenous glutamine synthetase (GS) enzyme. Since glutamine is critical for cell survival and growth, only cells that successfully incorporate the expression plasmid, which carries both the GS gene and the gene of interest (GOI), can survive in glutamine-free media. This method allows for the rapid identification of high-producing clones, significantly shortening timelines and increasing product titers.[40]

A recent advancement in this field is the transition from random integration toward semi-targeted systems using transposases, or to targeted integration (TI), by directing the GOI to "hotspots", transcriptionally active regions of the genome. High expression levels and greater genetic stability can be achieved.[41,42] These strategies have demonstrated the capability to increase product titers and significantly reduce the time and resources required for cell line development.

To ensure product consistency, regulatory agencies require that all cell banks used for the production of therapeutic proteins originate from a single cell. Therefore, single-cell cloning

DOI: 10.36838/IJHSR819.12

and imaging have become industry standard practices to follow the guidance and provide documented evidence of clonality.[43] Once the desired product-specific clone is identified, it is used to generate a GMP-compliant Master Cell Bank (MCB) and a subsequent Working Cell Bank (WCB). These banks must undergo rigorous qualification testing, such as identity verification, genetic stability analysis, and comprehensive safety assessments, including sterility, mycoplasma, and adventitious virus testing. GMP cell banks are typically stored long-term in the vapor phase of a liquid nitrogen tank to maintain maximum stability.

Cell Culture: The manufacturing process begins with thawing a vial from a stored cell bank, followed by a seed train using a series of culture vessels with increasing size to expand the biomass. Finally, the cultures are transferred to a large-scale stirred-tank bioreactor for protein production. To achieve high titers and desired product quality, both cell culture conditions and media components are studied and optimized. For bioreactor operations, process parameters such as agitation speed, temperature, dissolved oxygen (DO) levels, pH setpoints and bandwidth, and cultivation duration are controlled.[44] In addition, the dissolved CO_2 level (pCO_2) can significantly impact cell culture performance, particularly during scale-up, where gas mass transfer becomes more complex.[45] Therefore, the pCO_2 levels in the bioreactor are closely monitored and adjusted as needed by optimizing aeration and agitation strategies.

Cell culture media provide the essential nutrients, such as amino acids, glucose, vitamins, and trace metals, required for cell growth and high-quality protein production. The exact composition and concentration of these components are critical. Even minor changes can significantly influence product quality attributes, such as the glycosylation patterns on the protein.[46] These attributes may have an impact on the safety and efficacy of the product for patients. The majority of industrial mammalian cell culture processes are operated in fed-batch mode. In this approach, concentrated feed media are added periodically to the bioreactor while a lean basal media is used at the start of the culture. This strategy prevents nutrient depletion and extends the culture's longevity, ultimately leading to significantly higher product titers compared to simple batch processes. Nowadays, chemically defined and protein-free basal and feed media are commonly used to enhance performance and batch-to-batch consistency.[47]

Downstream: Following cell culture production, the secreted protein in the culture supernatant is harvested by centrifugation or membrane separation, and then purified through a sequence of downstream unit operations.[48] This process typically begins with affinity chromatography with high selectivity for primary capture, such as Protein A chromatography for monoclonal antibodies. To further refine the product, subsequent polishing steps, such as ion-exchange (IEX) or hydrophobic interaction chromatography (HIC), are incorporated to remove host cell proteins (HCPs), residual DNA, high molecular weight (HMW) aggregates, and other impurities.[49] These chromatography steps are operated using either bind-and-elute or flow-through modes, depending on the specific impurity profile and the intended purpose. Process parameters defining these chromatography unit operations typically include dynamic binding capacity, residence time, flow rate, buffer pH, and conductivity, and gradient slopes.[50,51] The overall objective of downstream purification is to achieve the target purity and quality attributes while maximizing overall yield.

In addition, the downstream process must incorporate an effective viral clearance strategy to reduce risks from both endogenous retrovirus-like particles and potential adventitious viral contaminants, ensuring the safety of therapeutic proteins derived from a biologics system. The use of multiple unit operations that employ orthogonal mechanisms of action based on distinct physical or chemical principles is required.[52] These operations typically include viral inactivation, such as low pH incubation or detergent treatment, and viral removal, such as nanofiltration and chromatography. The effectiveness of these steps is quantified by the Log Reduction Value (LRV). The cumulative LRV across the entire downstream process must demonstrate a substantial safety margin for both enveloped and non-enveloped viruses, covering a broad range of physical sizes and biochemical characteristics.[53]

At the end of the manufacturing process, the purified protein material is concentrated and exchanged into the defined formulation buffer using ultrafiltration/diafiltration (UF/DF). The formulation is established based on development studies demonstrating long-term stability under storage conditions.[54]

The typical manufacturing process for therapeutic proteins with the fed-batch cell culture mode takes 6-8 weeks from vial thaw to completion. Industrial cell culture production bioreactors range from 2,000L to 15,000L. Although specific process definitions vary by product, the fundamental unit operations remain highly similar, enabling the efficient use of existing GMP facilities and common equipment through a platform-based approach.

In recent years, an increasing number of bispecific antibodies are entering clinical trials for potential treatment for SLE, such as amulirafusp alfa, A-319, and teclistamab. Compared to standard monoclonal antibodies, bispecific formats exhibit significant structural diversity. Due to the inherent complexity of assembling two distinct antigen-binding functions within a single molecule, product-related impurities present a significant hurdle for process development. Specifically, downstream purification is more challenging and often requires comprehensive chromatographic strategies to effectively separate assembled bispecific molecules from impurities such as homodimers, half-antibodies, and other misassembled species.[55] Therefore, manufacturing considerations should be integrated into molecular design at an early stage to minimize the formation of challenging impurities.[56]

Dapirolizumab pegol, mentioned above, uses PEGylation technology in its molecule design. From the manufacturing point of view, PEGylation adds extra unit operations to the workflow. The process parameters, such as PEG reagent purity, molecular weight distribution, and reaction stoichiometry and duration, as well as pH and temperature, are controlled to obtain the desired conjugation efficiency and manage the degree of product heterogeneity.[57] Furthermore, specialized down-

stream purification steps are developed to separate the target conjugate from unconjugated PEG, sub-optimal PEGylated species, and other process-related byproducts. These steps are essential for achieving the high level of molecular homogeneity and purity demanded by regulatory standards.[58]

Overall, the manufacturing processes of therapeutic proteins for SLE follow established industrial platforms developed over the last 30 years. Utilizing these standardized workflows provides a clear, well-defined path through late-stage development, regulatory approval, and ultimately commercial usage.

Cell Therapy – Clinical Advances:

Cell therapy has been extensively investigated and successfully applied in oncology, particularly in hematologic malignancies, over the past 15–20 years.[59] Until recent years, its application in severe autoimmune diseases, including SLE, has begun to emerge.[60] Early clinical experiences, primarily from case reports in patients with refractory disease, have demonstrated encouraging efficacy and durable immunologic responses. These initial results have sparked a growing wave of investigations into diverse cell-based therapeutic strategies.[61] In this section, we highlight several of the main cell therapy modalities currently under evaluation in this field.

Autologous CAR-T cells: The landmark report of the first successful autologous CAR-T cell therapy for SLE emerged in 2021.[62] In this case, autologous CD19-directed CAR-T cells were used to treat a patient with severe, refractory disease, resulting in profound B-cell depletion, normalization of autoantibody and complement levels, and encouraging clinical improvement within weeks. The therapy was generally well tolerated with manageable acute toxicities. This work provided proof of concept that CAR-T therapy can function as an immune reset in autoimmune disease and laid the foundation for subsequent studies. Follow-up cases published later expanded on this initial finding by reporting treatment of multiple patients with refractory SLE.[63] Meanwhile, multiple CD19-directed autologous CAR-T candidates have been rapidly developed by other groups and have advanced into clinical trials, including CC-97540, Relma-cel, and CABA-201 (Rese-cel).[64-66] This therapeutic approach is also being explored in pediatric patients with remarkable efficacy.[67]

In addition, dual-target CAR-T strategies are being developed to achieve broader and more durable immune reset in SLE by simultaneously eliminating multiple pathogenic B-cell populations, which is similar to the bispecific antibody strategy. For example, AZD0120, a dual-target CAR-T therapy directed against CD19 and B cell maturation antigen (BCMA), is being evaluated.[68] Meanwhile, C-CAR168, a dual-target CAR-T therapy targeting CD20 and BCMA, is also under investigation in another clinical study.[69] Collectively, these dual-target CAR-T programs highlight a next-generation strategy designed to prolong immunologic remission by more comprehensively disrupting autoreactive B-cell populations.

On the other hand, the dual-CAR approach can also be achieved by a co-infusion strategy. Co-infusion of autologous CD19 and BCMA CAR-T cells is feasible and clinically active in patients with severe and refractory SLE.[70] It is worth noting that this strategy introduces additional manufacturing complexity, as it requires two separate processes for the production and release of two distinct CAR-T cell products, one for each target.

Allogeneic CAR-T cells: Another landmark study published in 2025 provides early clinical evidence that allogeneic anti-CD19 CAR-T cells can achieve meaningful disease control in patients with SLE.[71] Treatment with gene-engineered allogeneic CAR-T cells resulted in rapid and profound depletion of CD19-positive B cells, along with reductions in autoantibody titers and normalization of complement levels. Clinical remission was observed in several patients. Meanwhile, side effects consistent with known adverse events related to CAR-T were observed and manageable.[71] These data provide early proof-of-concept that allogeneic CAR-T therapy may offer a scalable option other than autologous approaches for immune reset in SLE. On the other hand, further studies are needed to understand response durability, immune rejection risk, and long-term safety.

Most recently, a phase 1 study of CB-010, a CRISPR-edited allogeneic anti-CD19 CAR-T cell therapy, was initiated in patients with refractory autoimmune diseases, including SLE, and has been granted Fast Track designation by the FDA.[72] BMS-986515 is another investigational, healthy donor–derived allogeneic CD19-targeted CAR-T cell therapy currently in phase 1 clinical development (NCT07115745).

Similar to autologous approaches, dual-target CAR-T strategies are also being explored in the allogeneic setting. For example, Allo-329 is an investigational allogeneic dual-target CAR-T cell therapy designed to simultaneously target CD19 and CD70.[73] CD70 is an inducible costimulatory ligand expressed on activated immune cells, including subsets of T cells and antigen-presenting cells. It is upregulated in inflammatory and autoimmune settings, including SLE. By targeting both antigens, Allo-329 aims to achieve broader immunomodulation and more effectively control immune escape or incomplete immune reset compared to single-target CAR-T therapies.

CAR-NK cells: While T-cell-based cell therapies remain the dominant approach, natural killer (NK) cells represent another viable option for CAR engineering and disease treatment. KN5501 is an investigational allogeneic anti-CD19 CAR-NK cell therapy being evaluated for the treatment of SLE.[74] In the ongoing phase 1 study, KN5501 was generally well tolerated, with low-grade CRS observed in some patients and no cases of other adverse events or infections. Most treated patients achieved meaningful disease improvement associated with rapid and complete depletion of CD19-positive B cells and normalization of immunologic markers.[74] These findings provide early clinical evidence that allogeneic CAR-NK cell therapy may represent a promising off-the-shelf alternative to autologous CAR-T approaches for immune reset in SLE. A similar study was conducted with LB0009 CAR-NK cells.[75] In addition, the dual-target CD19/BCMA CAR-NK (KN5601) is being tested in a clinical trial (NCT06792799).

DOI: 10.36838/IJHSR819.12

Cell Therapy – Manufacturing Process:

As a "living drug," the manufacturing process for cell therapies is inherently complex. However, since the approval of the first therapy in 2017, these processes have become more streamlined and better defined, particularly for autologous cell therapies. Their indications have also moved toward front-line treatment, benefiting a larger patient population. There are clear signs that this modality, including the manufacturing component, is becoming manageable.

For the autologous CAR-T cell therapy manufacturing process, it begins with leukapheresis, in which white blood cells are collected from the patient's blood and transported to a manufacturing facility. These cells are shipped either fresh or frozen under stringent logistics to ensure cell viability and sample traceability. At the facility, specific immune cells are isolated and then activated with stimulatory reagents to prepare them for genetic modification.[76] During the genetic engineering step, a viral vector, mainly a lentivirus or retrovirus vector, is used to introduce the genetic material that encodes a CAR into the cells, effectively "reprogramming" them to recognize and target specific markers such as CD19.[77] Then these genetically modified cells are expanded in culture vessels over several days to achieve a sufficient therapeutic dose. Following that, the cells are harvested, formulated, and cryopreserved as drug products. Finally, these drug products are returned to the clinical site for reinfusion into the patient after passing stringent quality-control testing to verify their identity, purity, and potency. In this scenario, each patient has specific CAR-T batch(s) originating from their own cells. The overall process time, called the "vein-to-vein" journey, could take 3–5 weeks.[78]

Next-generation autologous CAR-T manufacturing is shifting toward a shortened *ex vivo* cell culture timeline, often reduced from the traditional 9-14 days to as few as 24-48 hours, to improve product quality and patient outcomes.[79] By eliminating or significantly abbreviating the *ex vivo* expansion phase, manufacturers can produce "younger" T-cells that retain a less-differentiated, memory-like phenotype. These minimally cultured cells show enhanced stemness and lower levels of T-cell exhaustion, leading to superior *in vivo* expansion and greater anti-tumor potency compared to cells subjected to long-term culture.[80] This accelerated workflow not only addresses the risk of disease progression during the "vein-to-vein" waiting period but also simplifies the manufacturing process by reducing the need for larger culture capacity and expensive culture media. Several CAR-T programs using a rapid *ex vivo* manufacturing strategy are being evaluated in clinical trials, such as rapcabtagene autoleucel (YTB323) (NCT05798117).

Allogeneic cell therapy manufacturing represents a paradigm shift from the personalized "scale-out" approach of autologous therapies to a standardized "scale-up" model. It relies on healthy donor–derived cells and a scalable production model rather than patient-specific manufacturing.[81] In allogeneic processes, cells are collected from selected healthy donors and are used as the starting material for multiple patients after extensive characterization. The manufacturing process typically includes genetic modification of donor cells followed by cell expansion and cryopreservation.[82] Because allogeneic therapies use donor cells, the risks of immune rejection and graft-versus-host disease (GvHD) must be mitigated through genetic modifications, such as deletion of human leukocyte antigen (HLA) molecules or T-cell receptors.[83] Logistically, the allogeneic model eliminates the extreme time sensitivity and complex tracking required for individual patient lots. This transition not only significantly reduces the cost per dose through economies of scale but also allows for more sophisticated genetic engineering when needed.

CAR-NK cell manufacturing involves several specific requirements and challenges that distinguish it from T-cell-based processes.[84] First, NK cells are more difficult to expand *ex vivo*, requiring carefully optimized cytokines in culture media, and in some cases, feeder cells have been used. It adds complexity to process control, including the removal of residual feeder cells.[85] Second, genetic modification of NK cells is less efficient than in T cells, as NK cells are relatively resistant to viral transduction and electroporation. Higher vector doses must be used, or alternative gene-delivery platforms are needed, to achieve the desired CAR percentage while maintaining cell viability and function.[86,87] Third, NK cells are sensitive to over-activation and exhaustion. Therefore, tight control of activation status during manufacturing is required to balance cytotoxic efficacy and premature apoptosis or functional decline. In addition, cryopreservation and thaw recovery present unique challenges for NK cells, as post-thaw viability and functional activity can be more variable than for CAR-T products.[88] Given the challenges, scientists and researchers have established various ways to optimize the process, enabling CAR-NK cell therapies to advance to clinical trials.[89,90]

The integration of automation into autologous cell therapy manufacturing is essential for transitioning from manual, labor-intensive processes to a scalable, robust industrial model.[91] Currently, most of the cell therapy manufacturing workflows involve open-system manual interventions in many steps, which increases the risk of human error and contamination. By implementing closed and automated systems such as integrated modular robotic units, greater consistency can be achieved while significantly reducing the cleanroom footprint and addressing capacity challenges. These automated platforms often incorporate real-time monitoring and advanced sensors to capture process parameters so that better control over the complex biological variability coming from patient-derived starting materials can be achieved.[92] Ultimately, shifting toward an automated component can also reduce manufacturing and release resources and timelines, making these lifesaving "living drugs" more accessible to a larger patient population. Tremendous efforts are being made in this area to dramatically change the landscape of autologous cell therapy manufacturing. This strategy has the potential to become mainstream in the future once compliance with regulatory requirements is demonstrated and robust manufacturing processes are established.

mRNA-LNP – Clinical Advances:

Another emerging and rapidly advancing modality for SLE treatment is mRNA–LNP. Inspired by the strong efficacy of conventional CAR-T therapies, the mRNA-LNP approach

similarly aims to redirect T cells to eliminate pathogenic immune cell populations, such as autoreactive B cells.[93,94] However, important differences exist between the two strategies. Traditional CAR-T cell therapies, which rely mostly on viral vector-mediated gene transfer and *ex vivo* T-cell expansion, produce long-lived and proliferative CAR-T cells. In contrast, mRNA-LNP technology generates CAR-T cells directly within the patient, so-called *in vivo* CAR-T. Because mRNA expression is transient, which only lasts for several days and does not trigger genomic integration, it provides a potentially safer and more controllable method for modulating the immune system in chronic conditions such as SLE. Different from the systems utilized as COVID-19 vaccines, the application of mRNA-LNP in SLE treatment requires a high level of specificity through active delivery mechanisms. This can be achieved by decorating the nanoparticle surface with T cell targeting ligands to create a targeted nanoparticle (TNP), which can selectively engage T cells rather than relying on passive tissue distribution.[95]

Given the great potential of mRNA-LNP platforms in autoimmune diseases, numerous efforts are underway to advance these approaches into clinical trials for systematic evaluation of safety and efficacy. Among these, HN2301 represents the first wave of *in vivo* mRNA-LNP CAR-T candidates entering clinical trials with SLE patients.[96] HN2301 encapsulates CD19 CAR encoding mRNA within a proprietary T-cell targeted lipid nanoparticle (EnC-LNP), enabling direct reprogramming of $CD8^+$ T cells into functional anti-CD19 CAR-T cells following intravenous administration without the need for lymphodepleting chemotherapy. From a product design perspective, HN2301 incorporates a proprietary anti-CD8 antibody fragment for T-cell targeting and a proprietary ionizable lipid.[96]

According to early clinical data for HN2301, single or repeated administration at the 2 mg per infusion level induced measurable CAR-T cell formation and B-cell reduction. At the 4 mg per infusion level, up to 60% of circulating $CD8^+$ T cells were converted into CAR-expressing cells within six hours, resulting in complete depletion of circulating B cells that persisted for approximately 7-10 days.[96] Preliminary clinical observations also indicated meaningful reductions in SLEDAI-2K scores at three months post-infusion. HN2301 was generally well tolerated, with no grade ≥3 CRS, no neurotoxicity, and no other severe treatment-related adverse events reported.[96]

CPTX2309 is another investigational *in vivo* mRNA–LNP CAR-T therapy currently in Phase I clinical development for the treatment of SLE.[97] It also uses a CD8-targeted lipid nanoparticle, composed of a proprietary lipid and a humanized anti-CD8 targeting binder, to deliver anti-CD19 CAR-encoding mRNA directly into cytotoxic T cells, enabling them to express a CAR and deplete CD19+ B cells. Preclinical studies, including those conducted in non-human primates (NHP), demonstrated robust preferential engineering of $CD8^+$ CAR-T cells, rapid and profound B-cell depletion in blood and tissues, and subsequent repopulation with predominantly naïve B cells, suggestive of an immune reset similar to that observed with conventional CAR-T cell therapies.[97]

Compared to other therapeutic modalities, *in vivo* mRNA-LNP platforms for the treatment of SLE remain in the very early stages of clinical development. Nevertheless, the field is gaining significant momentum and attracting substantial interest from both academia and industry. It is driven by the platform's potential to achieve the immune reset observed in successful conventional CAR-T treatments, while offering a more convenient alternative that bypasses the complexities of *ex vivo* manufacturing of cell therapy. The question and potential hurdle for *in vivo* CAR technology is its capability of delivering the CAR gene to the right cells in the human body safely. Driving up transfection efficiency and specificity will be the key to lowering the required dose and expanding the safety window. Consequently, further NHP and human clinical trial data are essential to address these questions and gain a deeper understanding of the overall potential and relative standing of this approach compared with others.

mRNA-LNP – Manufacturing Process:

The manufacturing process of mRNA-LNP therapeutics begins with plasmid DNA (pDNA) production. First, an *E. coli* host strain is used to generate a product-specific cell bank, similar to a mammalian cell bank for protein production. A vial from that microbial cell bank is then thawed and expanded, followed by bioreactor fermentation to produce pDNA encoding the target mRNA sequence. The intracellular pDNA is then harvested through cell lysis, purified, and linearized via restriction enzyme digestion.[98] In addition to microbial-based production systems, cell-free DNA synthesis has also emerged in recent years as another option.[99]

Once the purified linear pDNA is available, mRNA can be manufactured through *in vitro* transcription (IVT) using it as a template. IVT is a rapid enzymatic reaction that synthesizes the desired mRNA in 2 to 3 hours.

Key process parameters for IVT include the concentration and purity of the pDNA template, T7 polymerase concentration and activity, nucleotide composition and ratios, reaction temperature, pH, magnesium ion concentration, and reaction duration. These parameters directly influence mRNA yield, integrity, and impurity levels, and therefore must be tightly controlled to ensure consistent and high-quality mRNA production. [100,101] Then, the synthesized mRNA is subsequently purified, typically using chromatography and tangential flow filtration, to remove impurities.

One of the critical impurities to control in mRNA-based products is double-stranded RNA (dsRNA), which is generated as a byproduct during the IVT process. dsRNA is a strong activator of innate immune receptors, leading to type I interferon production and pro-inflammatory signaling. dsRNA levels can be minimized through a holistic control strategy. It includes the optimization of IVT reaction conditions as well as plasmid template design and linearization process to minimize aberrant transcription.[102] In addition, the pDNA quality and T7 enzyme characteristics play an important role and can affect the dsRNA level in IVT.[103,104] Downstream purification

DOI: 10.36838/IJHSR819.12

steps, both chromatography and tangential flow filtration, can also be optimized to reduce residual dsRNA levels following IVT in the manufacturing process development.[105]

The LNP manufacturing process is based on the rapid and controlled mixing of an mRNA solution with lipids. Both impingement jet mixing (IJM) and microfluidic mixing technologies are commonly used in the industry.[106] During the manufacturing process, two solutions are prepared first: the first one is the mRNA solution in an acidic aqueous buffer, and the other is the lipid solution in ethanol containing four components (i.e., ionizable lipid, phospholipid, cholesterol, and PEG-lipid). Rapid mixing at the microscale induces spontaneous self-assembly of LNPs through electrostatic interactions between the protonated ionizable lipid and the negatively charged mRNA. Process parameters such as flow rate ratio, total flow rate, buffer pH, and temperature are tightly controlled to achieve reproducible product quality (e.g., particle size, narrow polydispersity, and high encapsulation efficiency).[106,107] After the encapsulation process, the crude LNP suspension has to go through a series of downstream processing steps such as buffer exchange, concentration, and sterile filtration.

The incorporation of targeting moieties to generate TNPs increases the complexity compared to the traditional LNP manufacturing process and presents unique challenges. For example, it requires conjugating T-cell targeting ligands to the nanoparticle surface. This step requires great control of ligand density and orientation to ensure consistent target function without negatively impacting particle integrity or mRNA encapsulation.[108] From a process standpoint, how to achieve consistent product quality becomes more challenging because surface modification can alter nanoparticle assembly and colloidal stability. Additional risks include increased aggregation and interference with sterile filtration.[109] Given the structural complexity of this type of mRNA-LNPs, advanced analytical tools and robust methods are important to understand the impact of manufacturing processes on product quality, ultimately enabling an effective control strategy that ensures batch-to-batch consistency.

Table 1: The Status of Different Biologics Modalities in Clinical Trials for SLE. The promising new biologic modalities are primarily in the early stages of clinical development, and their manufacturing processes are being further improved.

Modality	Approved drug (General)	Approved drug for SLE	Clinical Stage For SLE	Manufacturing Process
Therapeutic protein	Yes	Yes	Early and late stages	Well-established, large-scale production
Cell therapy – Autologous CAR-T	Yes	No	Early-stage	Established, one batch per patient
Cell therapy – Allogenic CAR-T	No	No	Early-stage	In development
Cell therapy – CAR-NK	No	No	Early-stage	In development
mRNA-LNP	Yes	No	Early-stage	Established for mRNA-LNP; In development for mRNA-TNP

Table 1 summarizes and compares the current status of each modality described above. With a strong interest in finding better treatments for SLE, it is no surprise to see other new modalities or new manufacturing strategies emerging. It is worth noting that although some initial and limited data from the new modalities are very promising, late-stage clinical trials with large patient populations are needed to demonstrate their true potential.

Conclusion

It is exciting to see that the SLE treatment landscape is rapidly advancing after many years of stagnation. The number of clinical trials is growing due to a better understanding of disease biology and a broad range of therapeutic strategies aimed at modulating various immune pathways. Different modalities, from therapeutic proteins to cell therapies to mRNA-LNP, are currently being explored in clinical trials to improve disease control. Importantly, early clinical signals from one modality can rapidly inform and inspire exploration of others. For example, soon after the promising results from autologous CAR-T cell therapies in SLE were published, other modalities, such as bispecific antibodies and mRNA-LNP-based approaches, were developed that seek to achieve similar clinical benefits through distinct and potentially more scalable technologies. Many of these approaches are still in early clinical development, waiting for more clinical data demonstrating overall benefits. On the other hand, differentiation has already extended beyond biological efficacy to the maturity of manufacturing platforms and supply-chain robustness, which are likely to become critical determinants of long-term benefit for patients. Overall, this review highlights the complexity of developing new treatments for patients and the importance of holistic consideration across separate but interconnected areas.

Acknowledgments

I want to thank my mentor, Dr. Trinny Tat, for her guidance. I also want to thank my parents for their support and encouragement.

References

1. Dai, X.; Fan, Y.; Zhao, X. Systemic Lupus Erythematosus: Updated Insights on the Pathogenesis, Diagnosis, Prevention and Therapeutics. *Signal Transduction and Targeted Therapy* **2025**, *10* (1), 102. https://doi.org/10.1038/s41392-025-02168-0.
2. Siegel, C. H.; Sammaritano, L. R. Systemic Lupus Erythematosus: A Review. *JAMA* **2024**, *331* (17), 1480–1491. https://doi.org/10.1001/jama.2024.2315.
3. Scopes, J.; Md Yusof, M. Y. Challenges in the Diagnosis of Systemic Lupus Erythematosus. *Clinical Medicine* **2025**, *25* (5), 100509. https://doi.org/10.1016/j.clinme.2025.100509.
4. Rahman Anisur; Isenberg David A. Systemic Lupus Erythematosus. *New England Journal of Medicine* **2008**, *358* (9), 929–939. https://doi.org/10.1056/NEJMra071297.
5. Aringer, M.; Costenbader, K.; Daikh, D.; Brinks, R.; Mosca, M.; Ramsey-Goldman, R.; Smolen, J. S.; Wofsy, D.; Boumpas, D. T.; Kamen, D. L.; Jayne, D.; Cervera, R.; Costedoat-Chalumeau, N.; Diamond, B.; Gladman, D. D.; Hahn, B.; Hiepe, F.; Jacobsen, S.; Khanna, D.; Lerstrøm, K.; Massarotti, E.; McCune, J.; Ruiz-Irastorza, G.; Sanchez-Guerrero, J.; Schneider, M.; Urowitz, M.; Bertsias, G.; Hoyer, B. F.; Leuchten, N.; Tani, C.; Tedeschi, S. K.; Touma, Z.; Schmajuk, G.; Anic, B.; Assan, F.; Chan, T. M.; Clarke, A. E.; Crow, M. K.; Czirják, L.; Doria, A.; Graninger, W.;

Halda-Kiss, B.; Hasni, S.; Izmirly, P. M.; Jung, M.; Kumánovics, G.; Mariette, X.; Padjen, I.; Pego-Reigosa, J. M.; Romero-Diaz, J.; Rúa-Figueroa Fernández, Í.; Seror, R.; Stummvoll, G. H.; Tanaka, Y.; Tektonidou, M. G.; Vasconcelos, C.; Vital, E. M.; Wallace, D. J.; Yavuz, S.; Meroni, P. L.; Fritzler, M. J.; Naden, R.; Dörner, T.; Johnson, S. R. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. *Annals of the Rheumatic Diseases* **2019**, *78* (9), 1151–1159. https://doi.org/10.1136/annrheumdis-2018-214819.

6. Barber, M. R. W.; Drenkard, C.; Falasinnu, T.; Hoi, A.; Mak, A.; Kow, N. Y.; Svenungsson, E.; Peterson, J.; Clarke, A. E.; Ramsey-Goldman, R. Global Epidemiology of Systemic Lupus Erythematosus. *Nature Reviews Rheumatology* **2021**, *17* (9), 515–532. https://doi.org/10.1038/s41584-021-00668-1.
7. Tsokos, George C. Systemic Lupus Erythematosus. *New England Journal of Medicine* **2011**, *365* (22), 2110–2121. https://doi.org/10.1056/NEJMra1100359.
8. Wallace, D. J. ANTIMALARIAL AGENTS AND LUPUS. *Rheumatic Disease Clinics of North America* **1994**, *20* (1), 243–263. https://doi.org/10.1016/S0889-857X(21)00233-7.
9. Furie, R.; Petri, M.; Zamani, O.; Cervera, R.; Wallace, D. J.; Tegzová, D.; Sanchez-Guerrero, J.; Schwarting, A.; Merrill, J. T.; Chatham, W. W.; Stohl, W.; Ginzler, E.M.; Hough, D. R.; Zhong, Z. J.; Freimuth, W.; van Vollenhoven, R. F.; BLISS-76 Study Group. A Phase III, Randomized, Placebo-Controlled Study of Belimumab, a Monoclonal Antibody That Inhibits B Lymphocyte Stimulator, in Patients with Systemic Lupus Erythematosus. *Arthritis & Rheumatism* **2011**, *63* (12), 3918–3930. https://doi.org/10.1002/art.30613.
10. Morand Eric F.; Furie Richard; Tanaka Yoshiya; Bruce Ian N.; Askanase Anca D.; Richez Christophe; Bae Sang-Cheol; Brohawn Philip Z.; Pineda Lilia; Berglind Anna; Tummala Raj. Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *New England Journal of Medicine* **2020**, *382* (3), 211–221. https://doi.org/10.1056/NEJMoa1912196.
11. Murphy, G.; Isenberg, D. A. New Therapies for Systemic Lupus Erythematosus — Past Imperfect, Future Tense. *Nature Reviews Rheumatology* **2019**, *15* (7), 403–412. https://doi.org/10.1038/s41584-019-0235-5.
12. Durcan, L.; O'Dwyer, T.; Petri, M. Management Strategies and Future Directions for Systemic Lupus Erythematosus in Adults. *The Lancet* **2019**, *393* (10188), 2332–2343. https://doi.org/10.1016/S0140-6736(19)30237-5.
13. Junt, T.; Calzascia, T.; Traggiai, E.; da Costa, A. N.; Gergely, P.; Schett, G.; Dörner, T.; Siegel, R. M. Defining Immune Reset: Achieving Sustained Remission in Autoimmune Diseases. *Nature Reviews Immunology* **2025**, *25* (7), 528–541. https://doi.org/10.1038/s41577-025-01141-w.
14. Mok, C. C. Targeted Small Molecules for Systemic Lupus Erythematosus: Drugs in the Pipeline. *Drugs* **2023**, *83* (6), 479–496. https://doi.org/10.1007/s40265-023-01856-x.
15. Blagov, A.; Orekhov, N.; Churov, A.; Starodubtseva, I.; Beloyartsev, D.; Kovyanova, T.; Sukhorukov, V.; Orekhov, A. Current Use and Development of Monoclonal Antibodies for the Treatment of Systemic Lupus Erythematosus: A Review. *Antibody Therapeutics* **2025**, *8* (1), 47–55. https://doi.org/10.1093/abt/tbae033.
16. Lyu, X.; Zhao, Q.; Hui, J.; Wang, T.; Lin, M.; Wang, K.; Zhang, J.; Shentu, J.; Dalby, P. A.; Zhang, H.; Liu, B. The Global Landscape of Approved Antibody Therapies. *Antibody Therapeutics* **2022**, *5* (4), 233–257. https://doi.org/10.1093/abt/tbac021.
17. Zhou, J.; Lei, B.; Shi, F.; Luo, X.; Wu, K.; Xu, Y.; Zhang, Y.; Liu, R.; Wang, H.; Zhou, J.; He, X. CAR T-Cell Therapy for Systemic Lupus Erythematosus: Current Status and Future Perspectives. *Frontiers in Immunology* **2024**, *Volume 15-2024*.
18. Chen, Z.; Pilehvar, E.; Sadeghi, H.; Pilehvar, Y. Precision Reimagined: CRISPR and Multiomics Transform Systemic Lupus Erythematosus Diagnosis and Therapy. *International Journal of Rheumatic Diseases* **2025**, *28* (4), e70189. https://doi.org/10.1111/1756-185X.70189.
19. Razavi, R.; Kegel, M.; Muscat-Rivera, J.; Weissman, D.; Melamed, J. R. Harnessing mRNA-Lipid Nanoparticles as Innovative Therapies for Autoimmune Diseases. *Molecular Therapy Methods & Clinical Development* **2025**, *33* (3), 101566. https://doi.org/10.1016/j.omtm.2025.101566.
20. Agmon-Levin, N.; Ignatenko, S.; Gordienko, A.; Cortés-Hernández, J.; Narongroeknawin, P.; Romanowska-Próchnicka, K.; Shen, N.; Ciferská, H.; Kodera, M.; Cheng-Chung, W. J.; Leszczynski, P.; Lan, J.-L.; Mysler, E.; Wojciechowski, R.; Tarr, T.; Vishneva, E.; Chen, Y.-H.; Kaneko, Y.; Finzel, S.; Hoi, A.; Okada, M.; Koolvisoot, A.; Lee, S.-S.; Lie, D.; Kaneko, H.; Rojkovich, B.; Sun, L.; Zotkin, E.; Viallard, J.-F.; López, B. P. M.; Ghanshani, S. A.; Lau, C.; Avrameas, A.; Sips, C.; Oliver, S. J. B Cell Depletion and BAFF Receptor Blockade with Ianalumab (VAY736) for the Treatment of Moderate-to-Severe Systemic Lupus Erythematosus: A Phase 2 Randomised, Double-Blind, Placebo-Controlled Trial with Subsequent Open-Label Treatment. *Annals of the Rheumatic Diseases* **2025**. https://doi.org/10.1016/j.ard.2025.11.015.
21. Pereira, N. A.; Chan, K. F.; Lin, P. C.; Song, Z. The "Less-Is-More" in Therapeutic Antibodies: Afucosylated Anti-Cancer Antibodies with Enhanced Antibody- Dependent Cellular Cytotoxicity. mAbs **2018**, *10* (5), 693–711. https://doi.org/10.1080/19420862.2018.1466767.
22. Furie Richard A.; van Vollenhoven Ronald F.; Kalunian Kenneth; Navarra Sandra; Romero-Diaz Juanita; Werth Victoria P.; Huang Xiaobi; Clark George; Carroll Hua; Meyers Adam; Musselli Cristina; Barbey Catherine; Franchimont Nathalie. Trial of Anti-BDCA2 Antibody Litifilimab for Systemic Lupus Erythematosus. *New England Journal of Medicine* **2022**, *387* (10), 894–904. https://doi.org/10.1056/NEJMoa2118025.
23. Furie, R.; Werth, V. P.; Milliman, E.; Ferber, K.; Brown, R.; Zoghbi, J.; Franchimont, N.; Lahoud, Y.; Barbey, C. Pharmacodynamic Effects of Litifilimab in Lupus in a Randomized, Placebo-Controlled Phase 2 Study: Rapid and Sustained Reductions in Type I Interferon-Associated Gene Expression and Cytokines. *Arthritis & Rheumatology* **2025**, *77* (12), 1726–1738. https://doi.org/10.1002/art.43271.
24. Furie, R. A.; Bruce, I. N.; Dörner, T.; Leon, M. G.; Leszczyński, P.; Urowitz, M.; Haier, B.; Jimenez, T.; Brittain, C.; Liu, J.; Barbey, C.; Stach, C. Phase 2, Randomized, Placebo-Controlled Trial of Dapirolizumab Pegol in Patients with Moderate-to-Severe Active Systemic Lupus Erythematosus. *Rheumatology* **2021**, *60* (11), 5397–5407. https://doi.org/10.1093/rheumatology/keab381.
25. Acharya, C.; Magnusson, M. O.; Vajjah, P.; Oliver, R.; Zamacona, M. Population Pharmacokinetics and Exposure-Response for Dapirolizumab Pegol From a Phase 2b Trial in Patients With Systemic Lupus Erythematosus. *The Journal of Clinical Pharmacology* **2023**, *63* (4), 435–444. https://doi.org/10.1002/jcph.2188.
26. Turecek, P. L.; Bossard, M. J.; Schoetens, F.; Ivens, I. A. PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs. *Journal of Pharmaceutical Sciences* **2016**, *105* (2), 460–475. https://doi.org/10.1016/j.xphs.2015.11.015.
27. Li, C.; Li, T.; Tian, X.; An, W.; Wang, Z.; Han, B.; Tao, H.; Wang, J.; Wang, X. Research Progress on the PEGylation of Therapeutic Proteins and Peptides (TPPs). *Frontiers in Pharmacology* **2024**, *Volume 15-2024*.

DOI: 10.36838/IJHSR819.12

28. Cai, J.; Gao, D.; Liu, D.; Liu, Z. Telitacicept for Autoimmune Nephropathy. *Frontiers in Immunology* **2023**, *Volume 14-2023*. https://doi.org/10.3389/fimmu.2023.1169084.
29. van Vollenhoven Ronald F.; Wang Li; Merrill Joan T.; Liu Yi; Bao Chunde; Li Fen; Hu Jiankang; Huang Chenghui; Zhao Jianhong; Huang Cibo; Mo Hanyou; Wei Wei; Lu Fu'ai; Li Jingyang; Zhao Dongbao; Wang Wenxiang; Li Lin; Zuraw Qing; Wang Xiaofei; Wang Xuebin; Fang Jianmin; Zhang Fengchun. A Phase 3 Trial of Telitacicept for Systemic Lupus Erythematosus. *New England Journal of Medicine* **2025**, *393* (15), 1475–1485. https://doi.org/10.1056/NEJMoa2414719.
30. Zhao, J.; Lin, C.; Xie, Q.; Shu, Q.; Cui, Y.; Luo, H.; Fan, W.; Huang, A.; Zhao, Y.; Fu, Z.; Xie, C.; Wu, H.; Yang, N.; He, L.; Feng, P.; Zhang, T.; Zhou, H.; Liu, W.; Hou, Q.; Mao, X.; Sun, J.; Chen, B.; Zeng, X. Anti-CD38 Monoclonal Antibody CM313 for Systemic Lupus Erythematosus: A Randomized, Double-Blind, Placebo-Controlled Phase Ib/IIa Trial. *Signal Transduction and Targeted Therapy* **2025**, *10* (1), 383. https://doi.org/10.1038/s41392-025-02487-2.
31. Yao, H.; Tian, W.; Zheng, Q.; Chen, M.; Jiang, G.; Liu, Z.; Nie, Y.; Wu, R.; Zheng, Z.; Li, Z. IMC-002 (IMM0306), a First-in-Class Bi-Specific Fusion Protein, Demonstrates Improvements in Systemic Lupus Erythematosus (SLE) Disease Activity Measures and Biomarkers in Patients with Moderate to Severe Active SLE in the Open-Label Phase 1b/2 Study. *Arthritis & Rheumatology* **2025**, *77* (Suppl 9).
32. Wang, S.; Peng, L.; Xu, W.; Zhou, Y.; Zhu, Z.; Kong, Y.; Leung, S.; Wang, J.; Yan, X.; Mi, J.-Q. Preclinical Characterization and Comparison between CD3/CD19 Bispecific and Novel CD3/CD19/CD20 Trispecific Antibodies against B-Cell Acute Lymphoblastic Leukemia: Targeted Immunotherapy for Acute Lymphoblastic Leukemia. *Frontiers of Medicine* **2022**, *16* (1), 139–149. https://doi.org/10.1007/s11684-021-0835-8.
33. Mei, C.; Guan, X.; Du, R.; Wu, B.; Song, Y.; Chen, X.; Li, M.; Liu, X.; Cheng, X.; Wang, W.; Wang, C.; Mei, H.; Duan, X.; Jiang, L.; Qiu, W.; Yu, L.; Liu, Y.; Wu, D.; Quan, G.; Chakhabi, B.; Zhao, X.; Zhong, X.; Xue, S.; Shen, W.; Tan, Y.; Yu, G.; Tu, G.; Chen, H.; Sun, A.; Liang, J.; Song, S.; Xu, J.; Yan, X.; Huang, A.; Li, Q. A-319, a CD3×CD19 T Cell Engager, for the Treatment of Refractory Systemic Lupus Erythematosus: Preliminary Evidence of Autoimmune Reset. *Arthritis & Rheumatology* **2025**, *77* (Suppl 9).
34. Moreau Philippe; Garfall Alfred L.; van de Donk Niels W.C.J.; Nahi Hareth; San- Miguel Jesús F.; Oriol Albert; Nooka Ajay K.; Martin Thomas; Rosinol Laura; Chari Ajai; Karlin Lionel; Benboubker Lotfi; Mateos Maria-Victoria; Bahlis Nizar; Popat Rakesh; Besemer Britta; Martínez-López Joaquín; Sidana Surbhi; Delforge Michel; Pei Lixia; Trancucci Danielle; Verona Raluca; Girgis Suzette; Lin Shun X.W.; Olyslager Yunsi; Jaffe Mindy; Uhlar Clarissa; Stephenson Tara; Van Rampelbergh Rian; Banerjee Arnob; Goldberg Jenna D.; Kobos Rachel; Krishnan Amrita; Usmani Saad Z. Teclistamab in Relapsed or Refractory Multiple Myeloma. *New England Journal of Medicine* **2022**, *387* (6), 495–505. https://doi.org/10.1056/NEJMoa2203478.
35. Alexander Tobias; Krönke Jan; Cheng Qingyu; Keller Ulrich; Krönke Gerhard. Teclistamab-Induced Remission in Refractory Systemic Lupus Erythematosus. *New England Journal of Medicine* **2024**, *391* (9), 864–866. https://doi.org/10.1056/NEJMc2407150.
36. Kang, X.; Zhao, Y.; Ling, H.; Huang, X. Bispecific and Multispecific T-Cell Engagers: Advancing the Future of Immunotherapy. *Antibody Therapeutics* **2025**, tbaf026. https://doi.org/10.1093/abt/tbaf026.
37. Wurm, F. M. Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells. *Nature Biotechnology* **2004**, *22* (11), 1393–1398. https://doi.org/10.1038/nbt1026.
38. Walsh, G.; Walsh, E. Biopharmaceutical Benchmarks 2022. *Nature Biotechnology* **2022**, *40* (12), 1722–1760. https://doi.org/10.1038/s41587-022-01582-x.
39. Fischer, S.; Handrick, R.; Otte, K. The Art of CHO Cell Engineering: A Comprehensive Retrospect and Future Perspectives. *Biotechnology Advances* **2015**, *33* (8), 1878–1896. https://doi.org/10.1016/j.biotechadv.2015.10.015.
40. Fan, L.; Kadura, I.; Krebs, L. E.; Hatfield, C. C.; Shaw, M. M.; Frye, C. C. Improving the Efficiency of CHO Cell Line Generation Using Glutamine Synthetase Gene Knockout Cells. *Biotechnology and Bioengineering* **2012**, *109* (4), 1007–1015. https://doi.org/10.1002/bit.24365.
41. Zeh, N.; Schmidt, M.; Schulz, P.; Fischer, S. The New Frontier in CHO Cell Line Development: From Random to Targeted Transgene Integration Technologies. *Biotechnology Advances* **2024**, *75*, 108402. https://doi.org/10.1016/j.biotechadv.2024.108402.
42. Schmieder, V.; Fieder, J.; Drerup, R.; Gutierrez, E. A.; Guelch, C.; Stolzenberger, J.; Stumbaum, M.; Mueller, V. S.; Higel, F.; Bergbauer, M.; Bornhoefft, K.; Wittner, M.; Gronemeyer, P.; Braig, C.; Huber, M.; Reisenauer-Schaupp, A.; Mueller, M. M.; Schuette, M.; Puengel, S.; Lindner, B.; Schmidt, M.; Schulz, P.; Fischer, S. Towards Maximum Acceleration of Monoclonal Antibody Development: Leveraging Transposase-Mediated Cell Line Generation to Enable GMP Manufacturing within 3 Months Using a Stable Pool. *Journal of Biotechnology* **2022**, *349*, 53–64. https://doi.org/10.1016/j.jbiotec.2022.03.010.
43. Frye, C.; Deshpande, R.; Estes, S.; Francissen, K.; Joly, J.; Lubiniecki, A.; Munro, T.; Russell, R.; Wang, T.; Anderson, K. Industry View on the Relative Importance of "Clonality" of Biopharmaceutical-Producing Cell Lines. *Biologicals* **2016**, *44* (2), 117–122. https://doi.org/10.1016/j.biologicals.2016.01.001.
44. Hu, W.-S.; Aunins, J. G. Large-Scale Mammalian Cell Culture. *Current Opinion in Biotechnology* **1997**, *8* (2), 148–153. https://doi.org/10.1016/S0958-1669(97)80093-6.
45. Matsunaga, N.; Kano, K.; Maki, Y.; Dobashi, T. Culture Scale-up Studies as Seen from the Viewpoint of Oxygen Supply and Dissolved Carbon Dioxide Stripping. *Journal of Bioscience and Bioengineering* **2009**, *107* (4), 412–418. https://doi.org/10.1016/j.jbiosc.2008.12.016.
46. Markert, S.; Torkler, S.; Hohmann, K.; Popp, O. Traces Matter: Targeted Optimization of Monoclonal Antibody N-Glycosylation Based on/by Implementing Automated High-Throughput Trace Element Screening. *Biotechnology Progress* **2020**, *36* (6), e3042. https://doi.org/10.1002/btpr.3042.
47. Ritacco, F. V.; Wu, Y.; Khetan, A. Cell Culture Media for Recombinant Protein Expression in Chinese Hamster Ovary (CHO) Cells: History, Key Components, and Optimization Strategies. *Biotechnology Progress* **2018**, *34* (6), 1407–1426. https://doi.org/10.1002/btpr.2706.
48. Shukla, A. A.; Hubbard, B.; Tressel, T.; Guhan, S.; Low, D. Downstream Processing of Monoclonal Antibodies—Application of Platform Approaches. *Journal of Chromatography B* **2007**, *848* (1), 28–39. https://doi.org/10.1016/j.jchromb.2006.09.026.
49. Liu, H. F.; Ma, J.; Winter, C.; Bayer, R. Recovery and Purification Process Development for Monoclonal Antibody Production. *mAbs* **2010**, *2* (5), 480–499. https://doi.org/10.4161/mabs.2.5.12645.
50. Jiang, C.; Flansburg, L.; Ghose, S.; Jorjorian, P.; Shukla, A. A. Defining Process Design Space for a Hydrophobic Interaction Chromatography (HIC) Purification Step: Application of Quality

by Design (QbD) Principles. *Biotechnology and Bioengineering* **2010**, *107* (6), 985–997. https://doi.org/10.1002/bit.22894.

51. Urmann, M.; Graalfs, H.; Joehnck, M.; Jacob, L. R.; Frech, C. Cation-Exchange Chromatography of Monoclonal Antibodies: Characterization of a Novel Stationary Phase Designed for Production-Scale Purification. *mAbs* **2010**, *2* (4), 395–404. https://doi.org/10.4161/mabs.12303.
52. Soni, D. H.; Reghellin, V.; Sbarufatti, G.; Minghetti, P.; Altomare, A. Viral Clearance in Biopharmaceutical Manufacturing: Current Strategies, Challenges, and Future Directions. *Biotechnology Advances* **2026**, *87*, 108784. https://doi.org/10.1016/j.biotechadv.2025.108784.
53. Miesegaes, G.; Lute, S.; Brorson, K. Analysis of Viral Clearance Unit Operations for Monoclonal Antibodies. *Biotechnology and Bioengineering* **2010**, *106* (2), 238–246. https://doi.org/10.1002/bit.22662.
54. Menzen, T.; Le Vay, K.; Arsiccio, A.; Helbig, C.; Hausmann, K.; Pabstmann, T.; Hawe, A. To Platform or Not to Platform: Strategic Considerations for Antibody Formulation in Early Clinical Development. *Journal of Pharmaceutical Sciences* **2026**, *115* (1), 104054. https://doi.org/10.1016/j.xphs.2025.104054.
55. Chen, S. W.; Zhang, W. Current Trends and Challenges in the Downstream Purification of Bispecific Antibodies. *Antibody Therapeutics* **2021**, *4* (2), 73–88. https://doi.org/10.1093/abt/tbab007.
56. Amash, A.; Volkers, G.; Farber, P.; Griffin, D.; Davison, K. S.; Goodman, A.; Tonikian, R.; Yamniuk, A.; Barnhart, B.; Jacobs, T. Developability Considerations for Bispecific and Multispecific Antibodies. *mAbs* **2024**, *16* (1), 2394229. https://doi.org/10.1080/19420862.2024.2394229.
57. Jevševar, S.; Kunstelj, M.; Porekar, V. G. PEGylation of Therapeutic Proteins. *Biotechnology Journal* **2010**, *5* (1), 113–128. https://doi.org/10.1002/biot.200900218.
58. Pfister, D.; Morbidelli, M. Process for Protein PEGylation. *Journal of Controlled Release* **2014**, *180*, 134–149. https://doi.org/10.1016/j.jconrel.2014.02.002.
59. Davis, S.; Lutfi, F. Advances in Chimeric Antigen Receptor T-Cell Therapy for Hematologic Malignancies and Autoimmune Conditions: A Review of the 2024 American Society of Hematology Annual Meeting. *Frontiers in Hematology* **2025**, *Volume 4-2025*.
60. Liu Yuxin; Dong Minghao; Chu Yunhui; Zhou Luoqi; You Yunfan; Pang Xiaowei; Yang Sheng; Zhang Luyang; Chen Lian; Zhu Lifang; Xiao Jun; Wang Wei; Qin Chuan; Tian Daishi; Gao Ting. Dawn of CAR-T Cell Therapy in Autoimmune Diseases. *Chinese Medical Journal* **2024**, *137* (10), 1140–1150. https://doi.org/10.1097/CM9.0000000000003111.
61. Wu, D.; Xu-Monette, Z. Y.; Zhou, J.; Yang, K.; Wang, X.; Fan, Y.; Young, K. H. CAR T-Cell Therapy in Autoimmune Diseases: A Promising Frontier on the Horizon. *Frontiers in Immunology* **2025**, *Volume 16-2025*.
62. Mougiakakos Dimitrios; Krönke Gerhard; Völkl Simon; Kretschmann Sascha; Aigner Michael; Kharboutli Soraya; Böltz Sebastian; Manger Bernhard; Mackensen Andreas; Schett Georg. CD19-Targeted CAR T Cells in Refractory Systemic Lupus Erythematosus. *New England Journal of Medicine* **2021**, *385* (6), 567–569. https://doi.org/10.1056/NEJMc2107725.
63. Mackensen, A.; Müller, F.; Mougiakakos, D.; Böltz, S.; Wilhelm, A.; Aigner, M.; Völkl, S.; Simon, D.; Kleyer, A.; Munoz, L.; Kretschmann, S.; Kharboutli, S.; Gary, R.; Reimann, H.; Rösler, W.; Uderhardt, S.; Bang, H.; Herrmann, M.; Ekici, A. B.; Buettner, C.; Habenicht, K. M.; Winkler, T. H.; Krönke, G.; Schett, G. Anti-CD19 CAR T Cell Therapy for Refractory Systemic Lupus Erythematosus. *Nature Medicine* **2022**, *28* (10), 2124–2132. https://doi.org/10.1038/s41591-022-02017-5.
64. Mueller, F.; Patel, K.; Reshef, R.; Cherry, M.; Nash, R.; Ayala, E.; Hsu, J.; Azzi, J.; Vandenberghe, P.; Houot, R.; von Tresckow, B.; Stifano, G.; Koegel, A.; Falcon, S.; Chaudhry, B.; Kelly, L.; Thorpe, J.; Frisbee, A.; Charab, S.; Das, S.; Ly, T.; Ogasawara, K.; Caimi, P. Tolerability, Efficacy, Pharmacokinetics, and Pharmacodynamics of BMS-986353 (CC-97540), a CD19-Directed Chimeric Antigen Receptor (CAR) T Cell Therapy Manufactured Using a Next-Generation Process, for Severe, Refractory Autoimmune Diseases: Updated Data from Ongoing Phase 1, Multicenter, Open-Label Studies. *Blood* **2024**, *144*, 2088. https://doi.org/10.1182/blood-2024-194525.
65. Wang, Y.; Yang, S.; Yu, Y.; Xia, P.; Xie, C.; Zhang, C.; Lu, L. Preliminary Safety and Efficacy of Relmacabtagene Autoleucel (Relma-Cel) in Adults with Moderately to Severely Active Systemic Lupus Erythematosus: A Phase I Dose-Escalation Study. *Journal of Autoimmunity* **2025**, *157*, 103489. https://doi.org/10.1016/j.jaut.2025.103489.
66. Sheikh, S.; Derebail, V.; Grovner, N.; Gulati, G.; Abedi, M.; Hogan, J.; White, Y.; Dicasoli, C.; Estremera, R.; Volkov, J.; Nunez, D.; Standanlick, J.; Werner, M.; Cicarelli, J.; Lam, Q.; Furmanak, T.; Nezhad, F.; Basu, S.; Tummala, R.; Chang, D. RESET SLE: CLINICAL TRIAL EVALUATING CABA-201, A FULLY HUMAN, AUTOLOGOUS 4-1BB ANTI-CD19 CAR T CELL THERAPY IN NONRENAL SLE AND LUPUS NEPHRITIS. *J Rheumatol* **2025**, *52* (Suppl 1), 27. https://doi.org/10.3899/jrheum.2025-0390.O029.
67. He, X.; Hu, B.; Zhang, Y.; Liu, F.; Li, Q.; Zheng, C.; Shen, J.; Yang, Z.; Wang, J.; Ma, D.; Qian, C.; Lu, M.; Mao, J. Treatment of Two Pediatric Patients with Refractory Systemic Lupus Erythematosus Using CD19-Targeted CAR T-Cells. *Autoimmunity Reviews* **2025**, *24* (1), 103692. https://doi.org/10.1016/j.autrev.2024.103692.
68. Wu, C.; Fu, Y.; Liu, J.; Ye, Y.; Wang, R.; Li, G.; Li, P.; Weng, L.; Shen, L.; Ye, S.; Fu, Q. OP0074 Preliminary Results of CD19/BCMA Dual-Targeting FasTCAR-T Cells GC012F (AZD0120) in Patients with Refractory Systemic Lupus Erythematosus-an Open-Label, Single-Arm Study. *Annals of the Rheumatic Diseases* **2025**, *84*, 65. https://doi.org/10.1016/j.ard.2025.05.096.
69. Shen, N.; Ding, H.; Li, W.; Shen, Y.; Zhang, C.; Ye, Y.; Wang, R.; Yang, S.; Wu, C.; Dai, D.; Zheng, C.; Qian, Y.; Luo, X.; Trinh, T.; Zhu, J.; Huang, J.; Yao, Y. CLINICAL IMPACT OF C-CAR168, A NOVEL ANTI-CD20/BCMA COMPOSITE AUTOLOGOUS CAR T THERAPY, IN REFRACTORY LUPUS NEPHRITIS. *J Rheumatol* **2025**, *52* (Suppl 1), 25. https://doi.org/10.3899/jrheum.2025-0390.O027.
70. Feng, J.; Huo, D.; Hong, R.; Jin, X.; Cao, H.; Shao, M.; Wen, R.; Zhang, Q.; Zhang, M.; Fu, S.; Wang, D.; Xu, H.; Wei, G.; Cui, J.; Huang, S.; Cui, D.; Chang, A. H.; Liu, Z.; Lu, L.; Lin, J.; Hu, Y.; Huang, H. Co-Infusion of CD19-Targeting and BCMA-Targeting CAR-T Cells for Treatment-Refractory Systemic Lupus Erythematosus: A Phase 1 Trial. *Nature Medicine* **2025**, *31* (11), 3725–3736. https://doi.org/10.1038/s41591-025-03937-8.
71. Wang, X.; Zhang, Y.; Wang, H.; Wu, X.; He, C.; Lin, S.; Pang, K.; Li, Y.; Chen, Y.; Tang, X.; Liu, X.; Wang, J.; Ye, S.; Yan, R.; Guan, T.; Dai, B.; Lu, J.; He, H.; Lin, L.; Lu, H.; Li, T.; Zhou, L.; Ye, L.; Zhao, J.; Liu, Y.; Ta, N.; Wu, J.; Cai, W.; Wan, Z.; Zhang, S.; Sun, R.; Zhao, X.; Wang, J.; Lin, Y.; Ning, B.; Zhao, Z.; Tang, X.; Du, J.; Mao, Z.; He, Y.; Zheng, H.; Sun, L.; Lin, X.; Xu, H. Allogeneic CD19-Targeting T Cells for Treatment-Refractory Systemic Lupus Erythematosus: A Phase 1 Trial. *Nature Medicine* **2025**, *31* (11), 3713–3724. https://doi.org/10.1038/s41591-025-03899-x.
72. Chatham, W.; Fiorenza, S.; Weinmann-Menke, J.; Sanchez-Pernaute, O.; Lafayette, R.; Askanase, A.; Gerber, M.; Alexander, L.; Mao, P.; Garner, E.; Kwong, G.; Fowler, T.; Kufeldt, H.; Kerfs, B.;

DOI: 10.36838/IJHSR819.12

Brown, A.; Holland, C.; Mcwilliams, K.; Shah, G.; Garrett, A.; Bryan, M.; Chu, L.; Alambra, E.; Skoble, J.; Kochy, T.; Nesheiwat, T.; Zudaire, E.; Portella, S. A PHASE 1, MULTICENTER, OPEN-LABEL STUDY OF CB-010, A NEXT-GENERATION CRISPR-EDITED ALLOGENEIC ANTI-CD19 CAR-T CELL THERAPY, IN PATIENTS WITH REFRACTORY SYSTEMIC LUPUS ERYTHEMATOSUS (GALLOP). *J Rheumatol* **2025**, *52* (Suppl 1), 147. https://doi.org/10.3899/jrheum.2025-0390.PV119.

73. Zhang, K.; Li, J.; Nguyen, D.; Tan, N.; Cheng, H.; Huang, D.; Gopalakrishnan, S.; Roberts, Z.; Sommer, C.; Lauron, E. Preclinical Evaluation of ALLO-329: Allogeneic CD19 CAR T Cells Expressing an Anti-Rejection CD70 CAR for the Treatment of Autoimmune Diseases [Abstract]. *Arthritis & Rheumatology* **2024**, *76* (Suppl 9).
74. Yu, Y.; Kong, R.; Xu, X.; Liu, S.; Chen, Q.; Li, X.; Sun, M.; Yang, J.; Zhao, D.; Gao, J. Allogenic CD19 CAR NK Cells Therapy in Refractory Systemic Lupus Erythematosus: An Open-Label, Single Arm, Prospective and Interventional Clinical Trial [Abstract]. *Arthritis & Rheumatology* **2024**, *76* (Suppl 9).
75. Gao, J.; Yu, Y.; Kong, R.; Xu, X.; Liu, S.; Chen, Q.; Li, X.; Wu, Y.; Xu, E.; Sun, M.; Yang, J.; Li, M.; Zhao, D. LB0009 ALLOGENIC CD19 CAR NK CELL THERAPY IN REFRACTORY SYSTEMIC LUPUS ERYTHEMATOSUS-A CASE SERIES STUDY. *Annals of the Rheumatic Diseases* **2025**, *84*, 321. https://doi.org/10.1016/j.ard.2025.05.396.
76. 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. *Journal of Experimental Medicine* **2024**, *221* (2), e20230903. https://doi.org/10.1084/jem.20230903.
77. Jadlowsky, J. K.; Hexner, E. O.; Marshall, A.; Grupp, S. A.; Frey, N. V.; Riley, J. L.; Veloso, E.; McConville, H.; Rogal, W.; Czuczman, C.; Hwang, W.-T.; Li, Y.; Leskowitz, R. M.; Farrelly, O.; Karar, J.; Christensen, S.; Barber-Rotenberg, J.; Gaymon, A.; Aronson, N.; Bernstein, W.; Melenhorst, J. J.; Roche, A. M.; Everett, J. K.; Zolnoski, S. A.; McFarland, A. G.; Reddy, S.; Petrichenko, A.; Cook, E. J.; Lee, C.; Gonzalez, V. E.; Alexander, K.; Kulikovskaya, I.; Ramírez-Fernández, Á.; Minehart, J. C.; Ruella, M.; Gill, S. I.; Schuster, S. J.; Cohen, A. D.; Garfall, A. L.; Shah, P. D.; Porter, D. L.; Maude, S. L.; Levine, B. L.; Siegel, D. L.; Chew, A.; McKenna, S.; Lledo, L.; Davis, M. M.; Plesa, G.; Herbst, F.; Stadtmauer, E. A.; Tebas, P.; DiNofia, A.; Haas, A.; Haas, N. B.; Myers, R.; O'Rourke, D. M.; Svoboda, J.; Tanyi, J. L.; Aplenc, R.; Jacobson, J. M.; Ko, A. H.; Cohen, R. B.; June, C. H.; Bushman, F. D.; Fraietta, J. A. Long-Term Safety of Lentiviral or Gammaretroviral Gene-Modified T Cell Therapies. *Nature Medicine* **2025**, *31* (4), 1134–1144. https://doi.org/10.1038/s41591-024-03478-6.
78. Locke, F. L.; Hu, Z.-H.; Siddiqi, T.; Jacobson, C. A.; Nikiforow, S.; Ahmed, S.; Miklos, D. B.; Lin, Y.; Lunning, M. A.; Hill, B. T.; Ghobadi, A.; Miao, H.; Shahani, S.; Spooner, C.; Fu, C.; Patel, A. R.; Xu, H.; Pasquini, M. C. Real-World Impact of Time from Leukapheresis to Infusion (Vein-to-Vein Time) in Patients with Relapsed or Refractory (r/r) Large B-Cell Lymphoma (LBCL) Treated with Axicabtagene Ciloleucel. *Blood* **2022**, *140* (Supplement 1), 7512–7515. https://doi.org/10.1182/blood-2022-155603.
79. Agliardi, G.; Dias, J.; Rampotas, A.; Garcia, J.; Roddie, C. Accelerating and Optimizing CAR T-Cell Manufacture to Deliver Better Patient Products. *The Lancet Hematology* **2025**, *12* (1), e57–e67. https://doi.org/10.1016/S2352-3026(24)00273-4.
80. Ghassemi, S.; Durgin, J. S.; Nunez-Cruz, S.; Patel, J.; Leferovich, J.; Pinzone, M.; Shen, F.; Cummins, K. D.; Plesa, G.; Cantu, V. A.; Reddy, S.; Bushman, F. D.; Gill, S. I.; O'Doherty, U.; O'Connor, R. S.; Milone, M. C. Rapid Manufacturing of Non-Activated Potent CAR T Cells. *Nature Biomedical Engineering* **2022**, *6* (2), 118–128. https://doi.org/10.1038/s41551-021-00842-6.
81. Lonez, C.; Breman, E. Allogeneic CAR-T Therapy Technologies: Has the Promise Been Met? *Cells* **2024**, *13* (2), 146. https://doi.org/10.3390/cells13020146.
82. Chen, S.; van den Brink, M. R. M. Allogeneic "Off-the-Shelf" CAR T Cells: Challenges and Advances. *Best Practice & Research Clinical Hematology* **2024**, *37* (3), 101566. https://doi.org/10.1016/j.beha.2024.101566.
83. Su, J.; Zeng, Y.; Song, Z.; Liu, Y.; Ou, K.; Wu, Y.; Huang, M.; Li, Y.; Tu, S. Genome-Edited Allogeneic CAR-T Cells: The Next Generation of Cancer Immunotherapies. *Journal of Hematology & Oncology* **2025**, *18* (1), 90. https://doi.org/10.1186/s13045-025-01745-8.
84. Khanal, S.; Bhattarai, N. A Scalable Protocol for *Ex Vivo* Production of CAR-Engineered Human NK Cells. *Methods and Protocols* **2025**, *8* (5), 102. https://doi.org/10.3390/mps8050102.
85. Lee, J.; Kim, M.; Jeong, H.; Noh, I.; Park, J. Recent Advances in Feeder-Free NK Cell Expansion as a Future Direction of Adoptive Cell Therapy. *Frontiers in Immunology* **2025**, *Volume 16-2025*.
86. Zhou, Z.; Chen, Y.; Ba, Y.; Xu, H.; Zuo, A.; Liu, S.; Zhang, Y.; Weng, S.; Ren, Y.; Luo, P.; Cheng, Q.; Zuo, L.; Zhu, S.; Zhou, X.; Zhang, C.; Chen, Y.; Han, X.; Pan, T.; Liu, Z. Revolutionising Cancer Immunotherapy: Advancements and Prospects in Non-Viral CAR-NK Cell Engineering. *Cell Proliferation* **2025**, *58* (4), e13791. https://doi.org/10.1111/cpr.13791.
87. Ingegnere, T.; Segain, B.; Cozzani, A.; Carlsten, M.; Mitra, S.; Gaggero, S. Optimizing CAR-NK Cell Transduction and Expansion: Leveraging Cytokine Modulation for Enhanced Performance. *Current Protocols* **2024**, *4* (11), e70040. https://doi.org/10.1002/cpz1.70040.
88. Xu, R.; Shi, X.; Huang, H.; Tan, W.-S.; Cai, H. Development of a Me2SO-Free Cryopreservation Medium and Its Long-Term Cryoprotection on the CAR-NK Cells. *Cryobiology* **2024**, *114*, 104835. https://doi.org/10.1016/j.cryobiol.2023.104835.
89. Wang, X.; Byrne, M. E.; Liu, C.; Ma, M. T.; Liu, D. Scalable Process Development of NK and CAR-NK Expansion in a Closed Bioreactor. *Frontiers in Immunology* **2024**, *Volume 15-2024*.
90. Pfefferle, A.; Contet, J.; Wong, K.; Chen, C.; Verhoeyen, E.; Slichter, C. K.; Schluns, K. S.; Cursons, J.; Berry, R.; Nikolic, I.; Rautela, J.; Huntington, N. D. Optimization of a Primary Human CAR-NK Cell Manufacturing Pipeline. *Clinical & Translational Immunology* **2024**, *13* (5), e1507. https://doi.org/10.1002/cti2.1507.
91. Melocchi, A.; Schmittlein, B.; Sadhu, S.; Nayak, S.; Lares, A.; Uboldi, M.; Zema, L.; di Robilant, B. N.; Feldman, S. A.; Esensten, J. H. Automated Manufacturing of Cell Therapies. *Journal of Controlled Release* **2025**, *381*, 113561. https://doi.org/10.1016/j.jconrel.2025.02.057.
92. Blache, U.; Popp, G.; Dünkel, A.; Koehl, U.; Fricke, S. Potential Solutions for Manufacture of CAR T Cells in Cancer Immunotherapy. *Nature Communications* **2022**, *13* (1), 5225. https://doi.org/10.1038/s41467-022-32866-0.
93. Bot, A.; Scharenberg, A.; Friedman, K.; Guey, L.; Hofmeister, R.; Andorko, J. I.; Klichinsky, M.; Neumann, F.; Shah, J. V.; Swayer, A. J.; Trudeau, K.; Weissman, D.; Stephan, M. T.; Buchholz, C. J.; June, C. H. *In Vivo* Chimeric Antigen Receptor (CAR)-T Cell Therapy. *Nature Reviews Drug Discovery* **2025**. https://doi.org/10.1038/s41573-025-01291-5.
94. Hunter, T. L.; Bao, Y.; Zhang, Y.; Matsuda, D.; Riener, R.; Wang, A.; Li, J. J.; Soldevila, F.; Chu, D. S. H.; Nguyen, D. P.; Yong, Q.-C.; Ross, B.; Nguyen, M.; Vestal, J.; Roberts, S.; Galvan, D.; Vega, J. B.; Jhung, D.; Butcher, M.; Nguyen, J.; Zhang, S.; Fernandez, C.; Chen, J.; Herrera, C.; Kuo, Y.; Pica, E. M.; Mondal, G.; Mammen,

A. L.; Scholler, J.; Tanis, S. P.; Sievers, S. A.; Frantz, A. M.; Adams, G. B.; Shawver, L.; Farzaneh-Far, R.; Rosenzweig, M.; Karmali, P. P.; Bot, A. I.; June, C. H.; Aghajanian, H. *In Vivo* CAR T Cell Generation to Treat Cancer and Autoimmune Disease. *Science* **2025**, *388* (6753), 1311–1317. https://doi.org/10.1126/science.ads8473.

95. Lemgart, V.; Richards, J.; Strauss, L.; Jiao, J.; Quido, B.; Powers, L.; Caron, A.; Shin, E.; Boesch, A.; Shah, D.; Le, T.; Kunden, R.; Stewart, S.; Nasimi, F.; Maiti, S.; Ali, M.; Lihoradova, O.; Griset, A.; Borges, C. M.; Kuhlman, W.; Lawendowski, C.; Nielsen, U.; Fantin, V.; Drummond, D.; Shaffer, D.; Sawyer, A. J. *In Vitro* and *In Vivo* Activity of a Lipid Nanoparticle System for the *In Vivo* Generation of CAR T Cells. *Blood* **2024**, *144*, 3421. https://doi.org/10.1182/blood-2024-207566.
96. Wang Qian; Xiao Ze Xiu; Zheng Xu; Wang Guihong; Yang Linlin; Shi Lina; Xiang Nan; Wang Xingbing; Zha Gao-Feng; Schett Georg; Chen Zhu. *In Vivo* CD19 CAR T-Cell Therapy for Refractory Systemic Lupus Erythematosus. *New England Journal of Medicine* **2025**, *393* (15), 1542–1544. https://doi.org/10.1056/NEJMc2509522.
97. Aghajanian, H.; Frantz, A. M.; Riener, R.; Wang, A.; Bao, Y.; Zhang, Y.; Matsuda, D.; Li, J. J.; Chu, D.; Hunter, T. L.; Soldevila-Casals, F.; Yong, Q.; Nguyen, M.; Sievers, S. A.; Nguyen, D. P.; Roberts, S.; Galvan, D.; Vega, J.; Butcher, M.; Zhang, S.; Flynn, S.; Kuo, Y.; Tanis, S. P.; Scholler, J.; Adams, G. B.; Rosenzweig, M.; Karmali, P.; June, C.; Bot, A. I. Design and Preclinical Development of a Novel *In Vivo* Chimeric Antigen Receptor (CAR) Product for B-Cell Involved Diseases. *Cytotherapy* **2025**, *27* (5, Supplement), S161. https://doi.org/10.1016/j.jcyt.2025.03.323.
98. Shankar, R.; Schmeer, M.; Schleef, M. Producing Plasmid DNA Template for Clinical Grade RNA Vaccine Manufacture. In *RNA Vaccines: Methods and Protocols*; Kramps, T., Ed.; Springer US: New York, NY, **2024**; pp 303–319. https://doi.org/10.1007/978-1-0716-3770-8_14.
99. Adie, T.; Orefo, I.; Kysh, D.; Kondas, K.; Thapa, S.; Extance, J.; Duncan, T.; Rothwell, P. J. dbDNATM: An Advanced Platform for Genetic Medicines. *Drug Discovery Today* **2022**, *27* (2), 374–377. https://doi.org/10.1016/j.drudis.2021.09.018.
100. Stover, N. M.; Ahmadi, S.; Rosenfeld, J.; Destro, F.; Myerson, A. S.; Braatz, R. D. Model-Based Optimization of Fed-Batch *In Vitro* Transcription. *ChemBioChem* **2025**, *26* (21), e202500485. https://doi.org/10.1002/cbic.202500485.
101. Boman, J.; Marušič, T.; Seravalli, T. V.; Skok, J.; Pettersson, F.; Nemec, K. Š.; Widmark, H.; Sekirnik, R. Quality by Design Approach to Improve Quality and Decrease Cost of *in Vitro* Transcription of mRNA Using Design of Experiments. *Biotechnology and Bioengineering* **2024**, *121* (11), 3415–3427. https://doi.org/10.1002/bit.28806.
102. Wolczyk, M.; Szymanski, J.; Trus, I.; Naz, Z.; Tame, T.; Bolembach, A.; Choudhury, N. R.; Kasztelan, K.; Rappsilber, J.; Dziembowski, A.; Michlewski, G. 5′ Terminal Nucleotide Determines the Immunogenicity of IVT RNAs. *Nucleic Acids Research* **2025**, *53* (3), gkae1252. https://doi.org/10.1093/nar/gkae1252.
103. Dousis, A.; Ravichandran, K.; Hobert, E. M.; Moore, M. J.; Rabideau, A. E. An Engineered T7 RNA Polymerase That Produces mRNA Free of Immunostimulatory Byproducts. *Nature Biotechnology* **2023**, *41* (4), 560–568. https://doi.org/10.1038/s41587-022-01525-6.
104. Martínez, J.; Lampaya, V.; Larraga, A.; Magallón, H.; Casabona, D. Purification of Linearized Template Plasmid DNA Decreases Double-Stranded RNA Formation during IVT Reaction. *Frontiers in Molecular Biosciences* **2023**, *Volume 10-2023*.
105. Siew, Y. Y.; Zhang, W. Removing Immunogenic Double-Stranded RNA Impurities Post *in Vitro* Transcription Synthesis for mRNA Therapeutics Production: A Review of Chromatography Strategies. *Journal of Chromatography A* **2025**, *1740*, 465576. https://doi.org/10.1016/j.chroma.2024.465576.
106. O'Brien Laramy, M. N.; Costa, A. P.; Cebrero, Y. M.; Joseph, J.; Sarode, A.; Zang, N.; Kim, L. J.; Hofmann, K.; Wang, S.; Goyon, A.; Koenig, S. G.; Hammel, M.; Hura, G. L. Process Robustness in Lipid Nanoparticle Production: A Comparison of Microfluidic and Turbulent Jet Mixing. *Mol. Pharmaceutics* **2023**, *20* (8), 4285–4296. https://doi.org/10.1021/acs.molpharmaceut.3c00390.
107. Hourdel, L.; Lebaz, N.; Peral, F.; Ripoll, M.; Briançon, S.; Bensaid, F.; Luthra, S.; Cogné, C. Overview on LNP-mRNA Encapsulation Unit Operation: Mixing Technologies, Scalability, and Influence of Formulation & Process Parameters on Physico-Chemical Characteristics. *International Journal of Pharmaceutics* **2025**, *672*, 125297. https://doi.org/10.1016/j.ijpharm.2025.125297.
108. Ko, J.; Tom, J.; Cheng, M. H. Y. Considerations on Antibody-Based Active Targeting Strategies for Lipid Nanoparticle Delivery Systems. *AAPS Open* **2026**, *12* (1), 19. https://doi.org/10.1186/s41120-026-00159-1.
109. Huang Ri; Wang Yifan; Shi Tianyu; Piao Xijun; Ren Keyun; Deng Changchang; Xu Congcong. Chemistry, Manufacturing, and Controls Strategies for Antibody-Targeted Lipid Nanoparticles. *BME Horiz* **2026**, *4* (2). https://doi.org/10.70401/bmeh.2026.0022.

Author

Phoebe Hu is a senior at the Academy for Allied Health Sciences. Having seen firsthand the impact of disease and illness, she is committed to pursuing research that contributes to advancements in treatment. She hopes to grow and apply herself to helping individuals live healthier, fuller lives.

DOI: 10.36838/IJHSR819.12

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Host Circadian Rhythm, Microbial Patterns, and Sleep Architecture

Poorvi Kurundkar

Oakridge International School, Gachibowli, Nanakramguda Road, Cyberabad, Hyderabad, Telangana, 500 008, India; kurundkar.poorvi@gmail.com

ABSTRACT: The human gut houses roughly 38 trillion microbes and contains about 1.3 times more microbial cells than the entire human body, making it an epicenter of revolutionary science. A growing body of research has demonstrated that intestinal microbiota not only modulates the human body's immune and neural systems, but also forms a close relation with circadian rhythm, which is predominantly regulated by the suprachiasmatic nucleus (the Central biological clock system) and peripheral clock systems. This paper aims to discuss the effects of feeding patterns, dietary interventions, and supplementation on sleep quality, thereby creating a bidirectional pathway governed by circadian rhythm. This provides a more effective way for shaping sleep structure and quality, including paradoxical sleep.

KEYWORDS: Biology, Chronobiology, Intestinal Clock Genes, Feeding Patterns, Sleep Disorders.

Introduction

Studying sleep disorders has become a physiological and psychological priority given their prevalence in today's population. Difficulty sleeping or insomnia symptoms are among the most common medical complaints affecting nearly a third of the adult population.[1] Short sleep has been linked with various adverse effects, such as mortality rate, diabetes mellitus, hypertension, cardiovascular disease, and obesity.[2] A Latin American Survey outlined that strategies patients used to cope with sleep problems included specific behaviors such as taking a warm bath, reading or watching TV (44 %), taking herbal beverages (17 %), or taking sleeping pills (10%).[3]

Although such practices have temporary symptomatic improvements due to increasing parasympathetic activity and reducing sleep onset latency, emerging evidence acknowledges that gut microbiota, particularly the interaction between circadian processes and microbial diversity, significantly affect sleep quality. This occurs through a bidirectional path, where circadian rhythmicity and sleep architecture function in symbiosis. Studies exploring the proliferation dynamics of gut microbiota are therefore essential not only for the regulation of sleep quality but also because sleep may modulate microbial balance itself.

Historically, mammalian physiology has adapted to periods in which mammals either forage and feed or sleep and fast.[4] The adaptation of physiology to these oscillations is controlled by a timekeeping system, which primarily encompasses the synchronization between light-dark signals from the eye with peripheral systems, including muscle, adipose tissue, liver, and intestine. It is also influenced by the gut-brain axis through the involvement of the neurotransmitters serotonin, melatonin, and the immune system. The key control for synchronizing external light cues with internal rhythm is the suprachiasmatic nucleus (SCN), sometimes referred to as the central circadian pacemaker, which is located in the hypothalamus and is responsible for this coordination of tissue-specific rhythms according to the light input it receives from the outside world.[5] The SCN utilizes a 24-hour pacemaking rhythm that is maintained by the cycling of a transcriptional-translational loop of CLOCK genes (Circadian Locomotor Output Cycles Kaput).

The clock gene mechanism is also present in cells throughout the brain and the body,[6] with genes such as Per1, Per2, and Bmal1 that regulate their own transcription by a negative feedback loop, generating cyclical inhibition of transcription factors. These act as timing signals in the gut, allowing certain microbial species to align their growth and metabolic patterns with the host's sleep-wake cycle.[7] In essence, the circadian rhythm in our bodies is directly associated with the light-dark cycle, which coordinates with metabolic rhythms that are primarily controlled by host signaling pathways and microbial activity in the gut.

Explorations into how sleep deprivation disrupts the immune system have blossomed into the discovery that it also promotes bacterial translocation and gut dysbiosis.[8] Today, it is well-established that a healthy gut microbiota is important for maintaining a healthy host and promoting optimal physiological function.[9] This will, in turn, aid in shaping the cyclical pattern of restorative sleep throughout a sleep period, which is known as sleep architecture. Sleep itself can be broken down into REM (rapid eye movement) and NREM (non-rapid eye movement) stages, which cyclically alternate throughout the night in a predictable structure known as sleep architecture, transitioning from lighter stages of NREM to deep slow-wave sleep before entering REM sleep.[10] The induction of many sleep disorders begins with the disorganization of this healthy structure and alternation of these stages.[11]

Many doctors have begun prescribing targeted probiotics for gut health as a way to manage certain sleep disorders. For example, according to a 2017 study, Lactobacillus gasseri CP2305 supplementation has improved the PSQI (Pittsburgh Sleep Quality Index) score by targeting stress-related insomnia.[12]

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.13

This growing trend fortifies connections between gut microbiota and sleep architecture, which this review will explore by evaluating patterns of feeding and dietary supplements with their effect on circadian rhythm.

More broadly, this review examines the ways in which gut microbiota influence circadian rhythm and sleep through interactions with intestinal clock genes and feeding patterns, thus depicting the importance of microbial synchronization for effective regulation of sleep.

Discussion

For more than half a century now, researchers have investigated the paradox that the brain's pacemaking center, despite occupying so little space, exerts influence over nearly every physiological process tied to timekeeping. Each cell in the SCN contains an autonomous molecular clock, and the SCN is composed of multiple of these single-cell oscillators, which can govern behavioral and physiological rhythmicity and be reset by environmental light. The key is that the heterogeneous network formed by the cellular clocks within the SCN must synchronize to maintain timekeeping activity.[13]

Host Pacemaking Mechanisms:

The discovery that the SCN contains tissue necessary for the generation of circadian rhythmicity was made through autoradiographic tracing methods in 1972.[14] The study found that specialized retinal ganglion cells (ipRGCs) in the eye contain melanopsin, a light-sensitive pigment that sends signals via the nerve fibers from the retina, which terminate at a specific region in the Hypothalamus- the SCN.

In the SCN, transcription factors are used to operate in a cyclical manner involving core clock genes (such as CLOCK, BMAL1, PER, and CRY) that regulate their own expression in a negative feedback loop.

The pioneering research that investigated the nature of this underlying feedback loop was first published in 1971 and used Drosophila Melanogaster (fruit fly) as the primary subject.[15] The method consisted of isolating the three distinct mutations that disrupted the flies' normal ~24-hour circadian rhythms: an arrhythmic mutant, a mutant with a shortened period of approximately 19 hours, and one with a lengthened period of approximately 28 hours. When these X-linked mutations were mapped in male offspring, all three different clock mutants always tracked to the same functional gene. This was termed the period "PER" gene, but it was still widely unknown how it worked until the late 1980s, when scientists discovered that PER protein levels cycle in ~24 h cycles and PER mRNA levels also cycle,[16] though delayed, suggesting a negative feedback mechanism.[17]

It turned out that CLOCK (CLK) forms a complex with CYCLE (CYC) and activates transcription of PER. Subsequently, PER, along with another protein, CRY, enters the nucleus and represses this transcription factor (CLK-CYC complex).[18] The mammalian functional equivalents of CLK and CYC are CLOCK and BMAL1, which form a heterodimer that binds to E-box elements in DNA, activating transcription of *PER* and *CRY* genes. This activation typically occurs in the morning, and repression (negative peak) occurs at night.[8]

Additionally, it is important to note that this is a two-loop feedback system, with core genes being *PER/CRY*, which control CLOCK expression; and secondary genes being *REV-ERB/RORα*, wherein *REV-ERBα* represses BMAL1 expression, and *RORα* activates BMAL1 expression.[8]

CLOCK and BMAL1 regulation in the SCN:

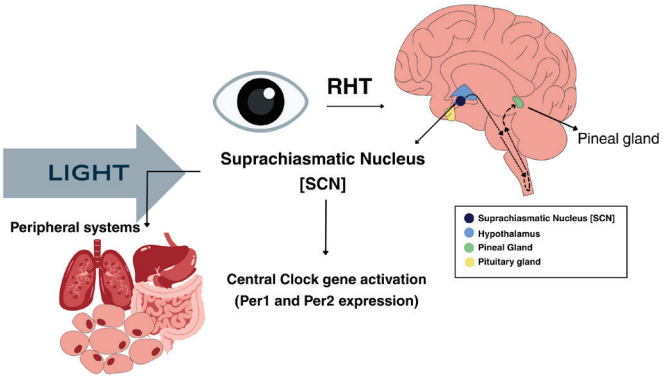


Figure 1: Light input from the retina reaches the SCN via the RHT, where it entrains central clock gene expression (***Per1, Per2***) and synchronizes circadian rhythm in peripheral bodily systems. Reproduced from: Matenchuk *et al.* 2020; 53:101340. doi:10.1016/j.smrv.2020.101340- Figure 2. Figure created with with canva.com

In order for the SCN to coordinate these peripheral molecular clocks via the RHT, external light cues must cause a rapid increase in PER1 and PER2 transcription in SCN neurons (see Figure 1), which causes a circadian clock reset each day.19

A 2002 study outlined that this aforementioned process requires neither rods nor cones, which were the only known retinal photoreceptors from previous publications. Instead, the study suggested the importance of specialized retinal ganglion cells that are intrinsically photosensitive to depolarization in response to light.[20] Thus, the retinal ganglion cells (ipRGCs) act as primary photoreceptors and fire action potentials via the RHT to the SCN through glutamate intersynaptic signaling upon light input to the retina. Within the SCN, glutamate binds to ionotropic glutamate receptors: AMPA Receptors (AMPARs) and NMDA Receptors (NMDARs). The by-product of this is a Ca^{2+} influx, which activates intracellular signaling cascades (such as CaMK and PKA), and promotes PER1 and PER2 transcription. Transcription of the *PER* gene will then repress the CLOCK: BMAL1 complex (which ties to the same cycle mentioned above), hence delaying PER1 and PER2 expression.

This process highlights the importance of timely light signals when it comes to circadian rhythm. Furthermore, since the *PER* gene expression oscillates across an approximate 24-hour period, light-induced changes or irregular light cues can alter the biological rhythm. This alteration is most apparent in conditions like jet lag, night-shift work, or chronic sleep disorders.[21] Such disruptions not only impair sleep and alertness but also extend to metabolic health and gut misalignment. For example, a 2018 study conducted among nursing staff revealed that the prevalence of metabolic syndrome (MetS: a cluster of

conditions that increase the risk of heart disease, stroke, and type II diabetes) in workers exposed to night shifts for up to 10 years was 37.5% higher when compared to diurnal workers.[22]

This establishes a clear relation between night work and metabolic syndrome as well as inflammatory dysfunction, which can be caused by gut dysbiosis (increased gut permeability), allowing bacterial components to translocate out of the gut and into the body, initiating low-grade systemic inflammation.[23]

CLOCK and BMAL1 regulation in the gut:

We have discussed how the SCN is a master regulator of circadian rhythm and acts to synchronize light exposure with peripheral clocks in the rest of the body, like the gut. Growing evidence pertaining to diurnal rhythmicity has revealed a strong bidirectional connection with the SCN and the microbiome situated in the gut, mediated through feeding-fasting cycles and microbial activity.

CLOCK and BMAL1 are expressed in intestinal epithelial cells (IECs), which was first discovered in 2007, when the effects of timed feeding and vagotomy were measured in mice.[24] The vagotomy, which consists of the removal of the vagus nerve, thus removing the connection between the gastrointestinal tract and SCN, showed that clock genes were expressed rhythmically throughout the gastrointestinal tract. Timed feeding shifted clock gene expression at the RNA and protein level but did not shift clock gene expression in the central clock. Thus, it is deduced that daytime feeding in nocturnal rodents is a strong synchronizer of gastrointestinal clock genes, but this synchronization occurs independently of the central clock. Furthermore, gastric clock gene expression is not mediated through the vagal nerve. The experiment proved the involvement of separate peripheral gastrointestinal clock genes, but their definite functions were not yet determined. However, subsequent research has revealed that this gastrointestinal clock was closely linked to feeding patterns, which act as a dominant external time cue, or zeitgeber, for the gut.[25]

Clock genes can affect a variety of functions in the gut. For example, BMAL1 and CLOCK increase expression of tight junction proteins, bolstering epithelial barrier function and allowing diurnal changes in intestinal permeability.[26] Expression of the nutrient transporters PEPT1 and SGLT1 is also controlled by clock gene oscillations. Furthermore, immune and inflammatory gene signatures are regulated by central transcription factors such as NF-κB, thus regulating inflammation according to feeding cycles.[27]

An example of this can be seen in a Circadian Rhythm Disruption (CRD) mouse model, along with induced downregulation of BMAL1, which was found to exhibit more severe colitis, accompanied by decreased tight junction protein levels and increased apoptosis of IECs compared with wild-type mice. This further suggests the tight link between clock genes like BMAL1, intestinal barrier function, and the expression of immune-related genes (apoptotic IEC's).[28]

Dietary Regimens and their influence on microbial rhythms:

Rhythmic food intake not only synchronizes the intestinal clock but also drives microbial oscillations, which illustrates the interdependence of host behavior and microbiota function within the diurnal framework.[29] A deeper dive into barrier integrity and permeability in the context of feeding patterns can facilitate a more complete understanding of this framework.

This rhythmic opening and tightening of the barrier may be adaptive: During the day, increased feeding promotes a controlled increase in permeability, allowing nutrient exchange and strengthening of immune barriers such as inflammatory cytokines. However, at night, tightening prevents inflammation and bacterial translocation.

Figure 2 captures information obtained from Scanning Electron Micrographs (SEMs) in mice,[30] which are nocturnal and have a dark phase (ZT12–ZT18) that corresponds to their active feeding period, whereas the light phase (ZT0–ZT6) is their resting or fasting period. Keeping this in mind, we can infer that during the feeding phase, the surfaces show dense bacterial adherence and microvilli crowding, as depicted in Figure 2. Along with this, the mucosal surface is coated with filamentous or rod-like microbial structures.

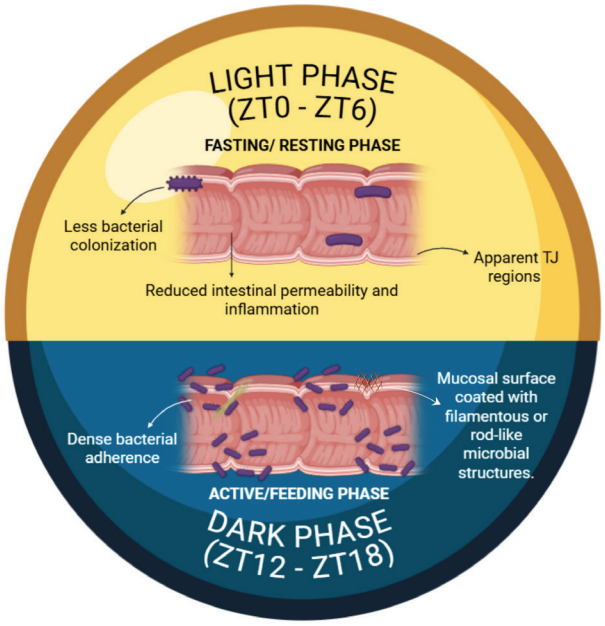


Figure 2: Model showing circadian variation in colonic epithelial adherence of mice. Bacterial attachment is markedly higher during the dark phase (ZT12–ZT18) compared to the light phase (ZT0–ZT6). Reproduced from Thaiss *et al.*, 2016 (Cell, 167(6), 1495–1510.e12)- Figure 2. Figure created with with BioRender.com

This indicates increased epithelial adherence and correlates with high nutrient availability due to feeding activity. During the fasting phase, the epithelial surface topography becomes visible and tight junctions are more apparent, appearing smoother and cleaner, with less bacterial colonization. This visually indicates less epithelial adherence, likely due to reduced nutrient availability.

Even if total bacterial numbers don't fluctuate dramatically, the genes expressed by microbes show evident oscillation. For example, carbohydrate metabolism enzymes (like glycosidases,

DOI: 10.36838/IJHSR819.13

polysaccharide-degrading enzymes) peak during feeding times when nutrient availability is highest. On the other hand, genes for flagellar motility (to move towards nutrient-rich areas) and host mucin degradation (enzymes like sialidases, fucosidases) peak during fasting times, when bacteria may need to move further for nutrients or consider other carbohydrate sources like mucin.[29] Another change is seen in the oscillation of SCFA (Short Chain Fatty Acid) production, which forms diurnal metabolites. These microbial metabolites then feed back into host circadian pathways by modulating histone acetylation, and thus regulating PER/ CRY/ CLOCK/ BMAL1 expression in gut epithelial cells.[30]

This proves how different food-intake patterns (feeding and fasting cycles) can affect bacterial populations in the gut. This rhythmic switching is exactly what intermittent fasting (IF) and time-restricted feeding (TRF) are based on. IF schemes have mainly two types, which have been applied to rodents, namely, daily fasting (DF) (e.g., 16:8, 20:4) and Alternate-Day Fasting (ADF). These methods restrict the timing of eating, not necessarily the content of the diet. IF triggers evolutionarily conserved cellular responses such as improved glucose regulation, stress resistance, and reduced inflammation by mimicking periods of scarcity (as in evolutionary history). A diet program of fasting every other day increased Firmicutes in male mice, while increasing *Lactobacillus* and *Bacteroides* in females. IF treatment in diabetic mice enriched beneficial bacteria like Lactobacillus and reduced harmful ones such as Akkermansia.[31]

On the other hand, High-Fat Diets (HFD) can increase the prevalence of metabolic diseases such as obesity, diabetes, non-alcoholic fatty liver disease, hypertension, and hyperuricemia. It is characterized by a significant intake of fats, typically comprising 60% or more of total daily calories, often with low carbohydrate intake. HFD-fed mice were observed to have a decrease in *Bacteroidetes* and an increase in Firmicutes (which break down complex carbohydrates into SCFAs), thus increasing the absorption of calories from food, contributing to obesity via adipose tissue inflammation.[32] Furthermore, an HFD has also been linked to disruptions in circadian rhythm, where mice on a regular chow diet maintained a normal free-running circadian rhythm (~23.6 hours), but consumption of an HFD lengthened their circadian period within just one week.[33] Interestingly enough, differences in mean body weight did not occur until the third week of HF feeding. Thus, the increase in period length was independent of body weight and overall activity levels.

While intermittent fasting affects the dynamics of when and how our gut microbes work by creating feeding windows, another daily behavior plays an important role in shaping the gut microbiome: sleep. Changes in sleep duration or quality can alter the balance of microbial species over the course of a day.

Dietary supplements and their impact on sleep and disorders:

Tononi and Cirelli's synaptic homeostasis hypothesis - "Sleep is the price the brain pays for plasticity"- suggests that disturbances in the brain's restorative processes can disrupt diurnal metabolic systems. Sleep is widely said to be directed by 2 processes: Sleep/wake homeostasis (regulating sleep needs based on the duration of wakefulness) and the circadian clock. The SCN acts as the primary circadian clock in this case, since it distinguishes between light-dark cues and also dictates hormonal fluctuations. For example, norepinephrine is present at higher levels in the morning, while melatonin and serotonin peak during the phase of darkness.[34] These hormonal timings help construct sleep's internal structure, which consists of alternating cycles of NREM and REM cycles. NREM is typically longer (75-80% of total sleep) and is categorized by parasympathetic activation, such as low blood pressure, low heart rate, and decreased temperature. REM, on the other hand, consists of shorter periods (20-25%), with a decrease in muscle tone, which is the continuous passive muscle contraction that provides resistance to passive stretch during a resting state.10 REM also causes greater sympathetic activation, such as increased heart rate, breathing, blood pressure, and temperature. REM sleep has more structured dreams, while NREM sleep has more unstructured "bizarre" dreams.[35]

Today, with a significant rise in irregular circadian timings, insomnia is affecting 11-16% of the global population.[36] It is characterized by decreased slow wave sleep due to reduced REM latency and increased REM density, implying that though the time taken to enter REM sleep is less, the number of rapid eye movements per minute is above average. This increased REM density is observed in eating disorders, narcolepsy, presenile dementia, and other neuropsychiatric diseases due to reduced or absent REM inhibition.[37] This imbalance of REM sleep stages has recently been linked to the impairment of intestinal permeability, causing a "leaky gut," through which bacteria and their end products have been found to translocate to the mesenteric lymph nodes, liver, and the brain. In Poroyko *et al.*,[38] mice were exposed to sleep fragmentation (SF) for 4 weeks, which caused increased food intake and reversible changes in gut microbiota composition, such as an increase in Lachnospiraceae and Ruminococcaceae and a decrease in Lactobacillaceae families. Excess Lachnospiraceae is known for impairment of glucose metabolism,[39] leading to inflammation, obesity, and promoting the onset of Type 1 diabetes,[40] depicting how sleep disruption can have vast effects on the gut.

However, the gut-brain axis can also impact sleep because a powerful innate sleep-inducing mechanism occurs via gut microbiota. Microbial activity, driven by feeding and fasting cycles, produces neuroactive metabolites such as Short-Chain Fatty Acids (SCFAs) or Gamma-aminobutyric acid (GABA). GABA acts as an inhibitory neurotransmitter that slows down neural signaling by acting as an anti-fight-or-flight chemical messenger.[41] Naturally produced from glutamate, GABA facilitates cross-barrier communication between the intestinal clock genes and central sleep-regulating centers, thereby affecting sleep onset latency and circadian regulation.[42] When this symbiotic relationship is disrupted through circadian misalignment or dietary regimens, it often manifests as sleep disorders such as insomnia or sleep fragmentation.[43] Thus, GABA can directly impact sleep quality as well as the GI tract. This connection has been evident in studies that revealed diminished levels of GABA in IBS-D patients as compared to controls.[44] GABA

also reduced the expression of proinflammatory cytokines, thus linking GI problems with disruptions in GABA signaling. Consequently, targeted dietary supplementation with probiotics or amino acids may serve as a strategy for restoring the structural integrity of sleep and mitigating disorders.

This means that exogenous supplementation may serve as a strategy to mend microbial diurnal oscillations that were broken due to jet lag, shift work, or sleep fragmentation. These interventions generally target the gut-brain axis through two distinct pathways: regulating the SCN and peripheral microbiome modification.

Primarily targeting the SCN, one of the most extensively researched and prescribed dietary supplements impacting sleep is Melatonin (MLT). MLT is a hormone naturally synthesized by the pineal gland, which acts via MT1 & MT2 receptors in the SCN. When it becomes dark, the SCN sends signals to the pineal gland to make melatonin, a hormone that promotes relaxation and reduced alertness by decreasing body temperature, thus aiding in sleep onset.[45] Patients with chronic insomnia were found to have significantly lower nighttime melatonin levels, with peak melatonin concentrations averaging 82.5 pg/mL, compared to 116.8 pg/mL in healthy sleepers.[46] MLT supplements are often utilized to manage sleep-related issues and have been purported to regulate sleep patterns and alleviate sleep disorders such as insomnia. More specifically, they're often prescribed to shift workers who have irregular sleep-wake schedules due to working at night, or patients with intense jet lag.[47]

Clinical trials have shown mixed results in humans, with a general decrease in sleep latency, but no consistent effect on sleep quality. Melatonin is known to be soporific, since it signals biological night to the SCN by telling it when to sleep; it is not restorative since it doesn't signal how much to sleep.

In animals, however, though melatonin is species & time-dependent, it can control architecture, such as reducing REM sleep after MLT during the dark period in nocturnal animals.[45] For example, Wang *et al.* outline that supplementation with melatonin ameliorates disturbances in the circadian rhythm homeostasis of the gut microbiota due to sleep restriction, by rescuing families like Muribaculaceae and Lachnospiraceae, which are sensitive to SF.[48] This suggests that supplementation can restore circadian rhythm as well as microbial diversity, and is likely why exogenous melatonin is commonly administered orally in cases of sleep disorders.[49]

Other than supplementation with melatonin itself, dietary strategies to enhance endogenous melatonin synthesis, particularly through tryptophan-rich foods, have been examined for their effects on sleep. According to a meta-analysis,[50] Tryptophan (Trp) supplementation has been found to shorten wake after sleep onset (-81.03 min/g).

Trp typically becomes Serotonin (see Figure 3), which converts to melatonin once triggered by darkness.[51]

Trp, being an essential amino acid, breaks down to form its metabolite, Tryptamine, which has been linked to the induction of melatonin synthesis in the past. In Lee *et al.*,[51] a study measuring melatonin synthesis in 10-week-old wild-type male C57BL/6 mice, Tryptamine was found to increase sleep duration to 31.6 ± 5.6 min compared to 18.6 ± 3.4 min in the control group. However, sleep onset latency and blood melatonin concentrations showed no significant increase for orally administered Tryptamine trials in the mice. This aligns with the aforementioned meta-analysis, which concluded that though Trp can increase sleep duration, Trp supplementation did not affect other sleep components like sleep efficiency or REM/NREM ratios. On the other hand, ITE, which is formed from the methylation of another metabolite of Trp (Indolokine A5), increased the expression of key genes for melatonin biosynthesis (Tph1a, Aadc, Aanat1, Aanat2, and Hiomt) in zebrafish embryos.[51]

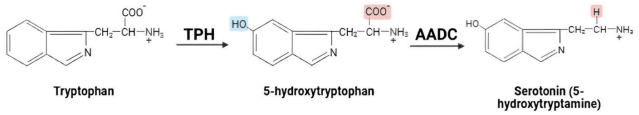


Figure 3: Biosynthetic pathway from L L-tryptophan to serotonin, which is subsequently converted to melatonin. ITE showed an incr ease in expression levels of key enzymes in this pathway. Reproduced from Lee *et al.*, ACS Omega, 2024, 9, 43875 43875–43883. Copyright 2024.

Emerging evidence suggests that other amino acids, such as phenylalanine, may exert even more pronounced effects on sleep. This is supported by Lee *et al.*, where phenylalanine's metabolite, phenethylamine, not only showed one of the longest sleep durations at 36.2 ± 3.4 min, but it also increased blood melatonin concentrations. The study thus concluded by restating phenethylamine's and ITE's potent role as a potential therapeutic agent for sleep disorders due to their role in melatonin synthesis. Despite these promising findings, it is important to remember that though melatonin in humans can decrease sleep latency and affect wake after sleep onset, there are much milder impacts on overall sleep efficiency and REM/NREM architecture.[45]

While melatonin and amino acid precursors primarily target central sleep architecture and neurotransmitter synthesis from the top down, an alternative strategy involves modulating the system from the bottom up via the gastrointestinal tract. This involves the employment of the gut-brain axis, specifically probiotics, for sleep regulation. Probiotics are living nutraceutical microorganisms, such as bacteria or yeast, that aid in digestion and supplement gastrointestinal flora.[52] Of the probiotics associated with sleep regulation, the Lactobacillus species have been of particular interest because they can produce neuroactive compounds that dictate sleep-wake rhythmicity.

Lactobacilli are characterized by their gram-positive rods and are the most common genus in the lactic acid bacteria (LAB) category.[53] LAB species can decarboxylate glutamate to produce γ-aminobutyric acid (GABA).

Though this production of GABA primarily occurs in the brain, the Enteric Nervous System (which governs the GI tract) is lined with GABA receptors, indicating GABA secreted by gut bacteria can be sensed locally by cells of the intestinal epithelial tract.[54] This is what connects Lactobacilli with sleep quality, as seen in Yu *et al.*, where milk was fermented with Lactobacillus brevis DL1-11, a strain with high GABA-producing capacity, and administered in different concentrations

DOI: 10.36838/IJHSR819.13

to mice (the highest concentration was termed high-dose GABA fermented milk (HGFM, 33.33 mg/kg.bw)).[55] The results indicated decreases in anxiety behavior after oral HGFM administration. Moreover, mice in the HGFM group exhibited a significantly prolonged sleep time and a shortened sleep latency. These results demonstrate a beneficial effect of Lactobacillus brevis on sleep. Additionally, significant increases in the relative abundances of Ruminococcus, Adlercreutzia, and Allobaculum, and the levels of certain SCFAs such as butyric acid were observed in the HGFM group.

This suggests that GABA-fermented milk and oral consumption of Lactobacillus strains with high GAD activity may improve sleep and regulate gut microbiota as well as SCFAs levels. The same is supported by various studies, such as one measuring sleep deprivation (SD) in Rhesus Monkeys, which speculated that SD-induced gut dysbiosis might be an outcome of reduced gut-derived GABA absorption.[56] The study thus stated that GABA-producing Lactobacillus supplementation might be beneficial for the treatment of SD-induced intestinal dysfunction.

Alongside Lactobacilli, Bifidobacteria, and similar psychobiotics can also influence sleep efficiency via stress modulation. They are gram-positive, nonmotile, and often branched anaerobic bacteria that are administered as dietary supplements for their probiotic properties.

A study measuring sleep efficiency (SE) found that higher SE, specifically by decreased Wake after sleep onset (WASO), is associated with higher levels of two Bifidobacterium species, *breve* and *saguini*.[57]

Targeting the gut microbiome as a therapeutic strategy for psychological disorders has shown promising outcomes in recent years.[58] This is why bifidobacteria are present in probiotics as well as psychobiotics. Species such as Bifidobacterium longum 1714 have been shown to influence stress responses and modulate neural responses during social stress, thereby impacting sleep quality.[59] In a study measuring eighty-nine adults with impaired sleep quality, B. longum 1714 significantly improved the PSQI component of sleep quality and daytime dysfunction due to sleepiness after 4 weeks and social functioning and energy/vitality after 8 weeks.[60] B. longum 1714 may modulate the HPA axis, and the SCFAs produced by the strain could influence melatonin synthesis, which emphasizes the multidirectional ability of certain probiotics to modulate neural activity, stress, mood, sleep and microbial diversity.

Overall, the evidence suggests that no single intervention fully restores circadian rhythm homeostasis, sleep architecture, and microbial rhythmicity simultaneously. Current evidence suggests that melatonin-based therapies possess the strongest translational capacity for circadian realignment, since melatonin supplementation remains the most clinically established intervention discussed. Unlike peripheral microbiome interventions, melatonin acts centrally, making it the most effective in disorders characterized primarily by circadian misalignment, such as jet lag, delayed sleep phase syndrome, or shift-work.[61] However, though the ability of exogenous melatonin to reduce sleep onset latency is consistently supported across human trials, recent studies have emerged suggesting that melatonin primarily functions as a chronobiotic rather than a restorative hypnotic agent.[45] This means that it can't efficiently improve total sleep efficiency, change NREM/REM ratios, or alter underlying sleep architecture in healthy adults compared to traditional hypnotic drugs such as benzodiazepines.

At the same time, the overall evidence supporting amino acid precursor therapies remains comparatively weak. Although certain metabolites increased sleep duration and melatonin concentrations in animal studies, many failed to significantly improve sleep onset latency, sleep efficiency, or REM/NREM architecture. Additionally, much of the current evidence derives from murine or zebrafish models, whose neuroendocrine regulation differs substantially from that of humans.[62] The mechanistic complexity of tryptophan metabolism also presents limitations, since tryptophan is diverted into multiple competing pathways, such as the kynurenine pathway.[63] Around 95% of tryptophan is processed via the kynurenine pathway, while only a small fraction (1%–2%) goes toward serotonin and melatonin synthesis,[64] which reduces the predictability of therapeutic outcomes.

Among microbiome-targeted therapies, psychobiotics such as Bifidobacterium longum 1714 may currently possess the greatest long-term promise because, unlike melatonin, which only affects the circadian rhythm, psychobiotics are able to affect the HPA axis, inflammation signaling, mood regulation, and neurotransmitters produced through SCFAs. This is particularly relevant because chronic insomnia in humans is frequently associated with hyperarousal, stress dysregulation, and anxiety rather than circadian misalignment alone. The observed improvements in PSQI scores, daytime dysfunction, energy levels, and social functioning following B. longum 1714 supplementation provides stronger human translational evidence than many other probiotic interventions. However, although psychobiotics may represent the future of sleep therapeutics, they are not yet as clinically standardized. Hence, further research should focus on optimizing such psychobiotic-based interventions through the identification of standardized dosages, treatment durations, and emphasis on specific probiotic strains. Greater emphasis will also likely be placed on determining how inter-individual gut microbiota variations influence therapeutic responsiveness,65 which could allow psychobiotic treatments to become more personalized and clinically reliable for the management of sleep disorders.

■ Conclusion

This paper primarily explored the interconnectedness between three mechanisms in the human body: gut microbiota, CLOCK gene expression, and sleep architecture, as seen in Figure 4.

Specifically, it discussed how gut-driven processes can affect sleep architecture. While central regulation of REM and NREM sleep is primarily governed by the brainstem and hypothalamic networks, peripheral signals originating from the gut can modulate sleep onset latency and quality. Microbiota-driven changes in GABAergic signaling, serotonergic–melatonergic pathways, and immune mediators are examples of mechanisms

by which IECs' clocks can shape sleep. This is further strengthened by research on dietary supplements, with compounds like exogenous Melatonin, Tryptophan, and Phenylalanine, which can induce sleep-like symptoms. The review highlights the importance of understanding how probiotic and psychobiotic supplementation, such as Lactobacillus and Bifidobacterium strains, can improve GABA secretion, thus supporting the idea that targeting the gut microbiota may offer a more holistic approach to circadian as well as sleep regulation compared to existing hormonal supplementation.

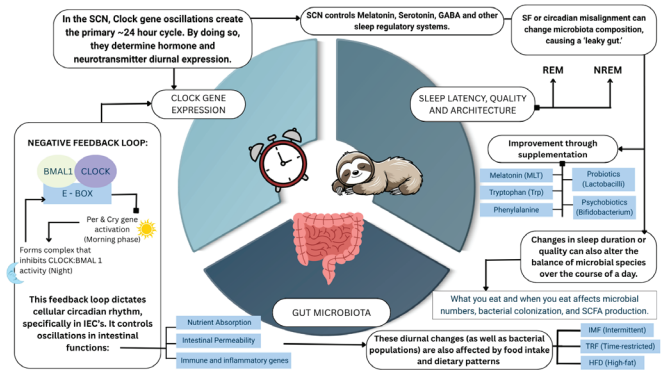


Figure 4: Summary of relationship between gut microbiota, sleep quality and clock gene expression.

Overall, this research highlights how gut microbiota and intestinal CLOCK genes interact with central neural pathways to influence sleep timing and architecture. These findings imply that optimization of sleep patterns requires targeting nutritional and microbiome-related factors rather than only focusing on central nervous system regulation. This perspective could enable the development of personalized chrononutrition strategies that focus on meal timing, dietary composition, and probiotic supplementation with an individual's endogenous circadian phase. This can be increasingly beneficial for overall human health because of the growing trends of sleep deprivation, shift work, and jet lag due to vast urbanization and globalization. Future research in this field should prioritize this kind of systems-level perspective, linking gut microbiota to sleep disorder therapies through well-controlled human studies and longitudinal monitoring studies, which have the potential to create interventions that can practically integrate chronobiological systems and the microbiome.

Acknowledgments

I sincerely thank Indigo Research as well as my mentor, Dr. Brydie R Huckestein (PhD. Microbiology and Immunology, University of Pittsburgh School of Medicine), for her invaluable guidance and support throughout the process of creating this review.

References

1. Karna, B.; Sankari, A.; Tatikonda, G. Sleep Disorder. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2026.
2. Itani, O.; Jike, M.; Watanabe, N.; Kaneita, Y. Short Sleep Duration and Health Outcomes: A Systematic Review, Meta-Analysis, and Meta-Regression. *Sleep Med.* **2017**, *32*, 246–256. https://doi.org/10.1016/j.sleep.2016.08.006.
3. Blanco, M.; Kriguer, N.; Lloret, S. P.; Cardinali, D. P. Attitudes towards Treatment among Patients Suffering from Sleep Disorders. A Latin American Survey. *BMC Fam. Pract.* **2003**, *4* (1). https://doi.org/10.1186/1471-2296-4-17.
4. Saini, C.; Suter, D. M.; Liani, A.; Gos, P.; Schibler, U. The Mammalian Circadian Timing System: Synchronization of Peripheral Clocks. *Cold Spring Harb. Symp. Quant. Biol.* **2011**, *76* (0), 39–47. https://doi.org/10.1101/sqb.2011.76.010918.
5. Reppert, S. M.; Weaver, D. R. Coordination of Circadian Timing in Mammals. *Nature* **2002**, *418* (6901), 935–941. https://doi.org/10.1038/nature00965.
6. Central Autonomic System. In *The Rat Nervous System*; Elsevier, **2015**; pp 629–673. https://doi.org/10.1016/b978-0-12-374245-2.00023-1.
7. Dunlap, J. C. Molecular Bases for Circadian Clocks. *Cell* **1999**, *96* (2), 271–290. https://doi.org/10.1016/s0092-8674(00)80566-8.
8. Matenchuk, B. A.; Mandhane, P. J.; Kozyrskyj, A. L. Sleep, Circadian Rhythm, and Gut Microbiota. *Sleep Med. Rev.* **2020**, *53*, 101340. https://doi.org/10.1016/j.smrv.2020.101340.
9. The Human Gut Microbiota. In *Advances in Experimental Medicine and Biology*; Springer International Publishing: Cham, **2016**; pp 95–108. https://doi.org/10.1007/978-3-319-31248-4_7.
10. Patel, A. K.; Reddy, V.; Shumway, K. R.; Araujo, J. F. Physiology, Sleep Stages. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2026.
11. Medic, G.; Wille, M.; Hemels, M. E. Short- and Long-Term Health Consequences of Sleep Disruption. *Nat. Sci. Sleep* **2017**, *9*, 151–161. https://doi.org/10.2147/NSS.S134864.
12. Nishida, K.; Sawada, D.; Kawai, T.; Kuwano, Y.; Fujiwara, S.; Rokutan, K. Para-psychobiotic Lactobacillus Gasseri CP 2305 Ameliorates Stress-related Symptoms and Sleep Quality. *J. Appl. Microbiol.* **2017**, *123* (6), 1561–1570. https://doi.org/10.1111/jam.13594.
13. Li, Y.; Liu, Z.; Zhang, J.; Wang, R.; Chen, L. Synchronization Mechanisms of Circadian Rhythms in the Suprachiasmatic Nucleus. *IET Syst. Biol.* **2009**, *3* (2), 100–112. https://doi.org/10.1049/iet-syb.2007.0057.
14. Weaver, D. R. The Suprachiasmatic Nucleus: A 25-Year Retrospective. *J. Biol. Rhythms* **1998**, *13* (2), 100–112. https://doi.org/10.1177/074873098128999952.
15. Konopka, R. J.; Benzer, S. Clock Mutants of Drosophila Melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **1971**, *68* (9), 2112–2116.
16. Hardin, P. E.; Hall, J. C.; Rosbash, M. Circadian Oscillations in Period Gene mRNA Levels Are Transcriptionally Regulated. *Proc. Natl. Acad. Sci.* **1992**, *89* (24), 11711–11715. https://doi.org/10.1073/pnas.89.24.11711.
17. Zehring, W. A.; Wheeler, D. A.; Reddy, P.; Konopka, R. J.; Kyriacou, C. P.; Rosbash, M.; Hall, J. C. P-Element Transformation with Period Locus DNA Restores Rhythmicity to Mutant, Arrhythmic Drosophila Melanogaster. *Cell* **1984**, *39* (2), 369–376. https://doi.org/10.1016/0092-8674(84)90015-1.
18. Vitaterna, M. H.; Shimomura, K.; Jiang, P. Genetics of Circadian Rhythms. *Neurol. Clin.* **2019**, *37* (3), 487–504. https://doi.org/10.1016/j.ncl.2019.05.002.
19. Ramkisoensing, A.; Meijer, J. H. Synchronization of Biological Clock Neurons by Light and Peripheral Feedback Systems Promotes Circadian Rhythms and Health. *Front. Neurol.* **2015**, *6*. https://doi.org/10.3389/fneur.2015.00128.
20. Berson, D. M.; Dunn, F. A.; Takao, M. Phototransduction by Retinal Ganglion Cells That Set the Circadian Clock. *Science* **2002**, *295* (5557), 1070–1073. https://doi.org/10.1126/science.1067262.

DOI: 10.36838/IJHSR819.13

21. Khan, S.; Nabi, G.; Yao, L.; Siddique, R.; Sajjad, W.; Kumar, S.; Duan, P.; Hou, H. Health Risks Associated with Genetic Alterations in the Internal Clock System by External Factors. *Int. J. Biol. Sci.* **2018**, *14* (7), 791–798. https://doi.org/10.7150/ijbs.23744.
22. Holanda, N. C. P. D.; De Castro Moreno, C. R.; Marqueze, E. C. Metabolic Syndrome Components: Is There a Difference According to Exposure to Night Work? *Chronobiol. Int*. **2018**, *35* (6), 801–810. https://doi.org/10.1080/07420528.2018.1450267.
23. Li, Y.; Hao, Y.; Fan, F.; Zhang, B. The Role of Microbiome in Insomnia, Circadian Disturbance, and Depression. *Front. Psychiatry* **2018**, *9*, 669. https://doi.org/10.3389/fpsyt.2018.00669.
24. Hoogerwerf, W. A.; Hellmich, H. L.; Cornélissen, G.; Halberg, F.; Shahinian, V. B.; Bostwick, J.; Savidge, T. C.; Cassone, V. M. Clock Gene Expression in the Murine Gastrointestinal Tract: Endogenous Rhythmicity and Effects of a Feeding Regimen. *Gastroenterology* **2007**, *133* (4), 1250–1260. https://doi.org/10.1053/j.gastro.2007.07.009.
25. Damiola, F.; Le Minh, N.; Preitner, N.; Kornmann, B.; Fleury-Olela, F.; Schibler, U. Restricted Feeding Uncouples Circadian Oscillators in Peripheral Tissues from the Central Pacemaker in the Suprachiasmatic Nucleus. *Genes Dev*. **2000**, *14* (23), 2950–2961. https://doi.org/10.1101/gad.183500.
26. Oh-oka, K.; Kono, H.; Ishimaru, K.; Miyake, K.; Kubota, T.; Ogawa, H.; Okumura, K.; Shibata, S.; Nakao, A. Expressions of Tight Junction Proteins Occludin and Claudin-1 Are under the Circadian Control in the Mouse Large Intestine: Implications in Intestinal Permeability and Susceptibility to Colitis. *PLoS ONE* **2014**, *9* (5), e98016. https://doi.org/10.1371/journal.pone.0098016.
27. Curtis, A. M.; Bellet, M. M.; Sassone-Corsi, P.; O'Neill, L. A. J. Circadian Clock Proteins and Immunity. *Immunity* **2014**, *40* (2), 178–186. https://doi.org/10.1016/j.immuni.2014.02.002.
28. Zhang, Z.; Li, W.; Han, X.; Tian, D.; Yan, W.; Liu, M.; Cao, L. Circadian Rhythm Disruption-Mediated Downregulation of Bmal1 Exacerbates DSS-Induced Colitis by Impairing Intestinal Barrier. *Front. Immunol.* **2024**, *15*, 1402395. https://doi.org/10.3389/fimmu.2024.1402395.
29. Thaiss, C. A.; Zeevi, D.; Levy, M.; Zilberman-Schapira, G.; Suez, J.; Tengeler, A. C.; Abramson, L.; Katz, M. N.; Korem, T.; Zmora, N.; Kuperman, Y.; Biton, I.; Gilad, S.; Harmelin, A.; Shapiro, H.; Halpern, Z.; Segal, E.; Elinav, E. Transkingdom Control of Microbiota Diurnal Oscillations Promotes Metabolic Homeostasis. *Cell* **2014**, *159* (3), 514–529. https://doi.org/10.1016/j.cell.2014.09.048.
30. Thaiss, C. A.; Levy, M.; Korem, T.; Dohnalová, L.; Shapiro, H.; Jaitin, D. A.; David, E.; Winter, D. R.; Gury-BenAri, M.; Tatirovsky, E.; Tuganbaev, T.; Federici, S.; Zmora, N.; Zeevi, D.; Dori-Bachash, M.; Pevsner-Fischer, M.; Kartvelishvily, E.; Brandis, A.; Harmelin, A.; Shibolet, O.; Halpern, Z.; Honda, K.; Amit, I.; Segal, E.; Elinav, E. Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. *Cell* **2016**, *167* (6), 1495-1510. e12. https://doi.org/10.1016/j.cell.2016.11.003.
31. Cai, Y.; Liu, Y.; Wu, Z.; Wang, J.; Zhang, X. Effects of Diet and Exercise on Circadian Rhythm: Role of Gut Microbiota in Immune and Metabolic Systems. *Nutrients* **2023**, *15* (12), 2743. https://doi.org/10.3390/nu15122743.
32. Mamun, M. A. A.; Rakib, A.; Mandal, M.; Singh, U. P. Impact of a High-Fat Diet on the Gut Microbiome: A Comprehensive Study of Microbial and Metabolite Shifts During Obesity. *Cells* **2025**, *14* (6), 463. https://doi.org/10.3390/cells14060463.
33. Kohsaka, A.; Laposky, A. D.; Ramsey, K. M.; Estrada, C.; Joshu, C.; Kobayashi, Y.; Turek, F. W.; Bass, J. High-Fat Diet Disrupts Behavioral and Molecular Circadian Rhythms in Mice. *Cell Metab*. **2007**, *6* (5), 414–421. https://doi.org/10.1016/j.cmet.2007.09.006.
34. Sánchez, S.; Sánchez, C.; Paredes, S. D.; Cubero, J.; Rodríguez, A. B.; Barriga, C. Circadian Variations of Serotonin in Plasma and Different Brain Regions of Rats. *Mol. Cell. Biochem.* **2008**, *317* (1–2), 105–111. https://doi.org/10.1007/s11010-008-9836-z.
35. Perogamvros, L.; Dang-Vu, T. T.; Desseilles, M.; Schwartz, S. Sleep and Dreaming Are for Important Matters. *Front. Psychol.* **2013**, *4*. https://doi.org/10.3389/fpsyg.2013.00474.
36. Chen, C. N. Depression, Antidepressants and Sleep. *Adv. Biosci.* **1978**, *21*, 213–216.
37. Gobbi, G.; Comai, S. Differential Function of Melatonin MT1 and MT2 Receptors in REM and NREM Sleep. *Front. Endocrinol*. **2019**, *10*, 87. https://doi.org/10.3389/fendo.2019.00087.
38. Poroyko, V. A.; Carreras, A.; Khalyfa, A.; Khalyfa, A. A.; Leone, V.; Peris, E.; Almendros, I.; Gileles-Hillel, A.; Qiao, Z.; Hubert, N.; Farré, R.; Chang, E. B.; Gozal, D. Chronic Sleep Disruption Alters Gut Microbiota, Induces Systemic and Adipose Tissue Inflammation and Insulin Resistance in Mice. *Sci. Rep.* **2016**, *6* (1), 35405. https://doi.org/10.1038/srep35405.
39. Vacca, M.; Celano, G.; Calabrese, F. M.; Portincasa, P.; Gobbetti, M.; De Angelis, M. The Controversial Role of Human Gut Lachnospiraceae. *Microorganisms* **2020**, *8* (4), 573. https://doi.org/10.3390/microorganisms8040573.
40. Kostic, A. D.; Gevers, D.; Siljander, H.; Vatanen, T.; Hyötyläinen, T.; Hämäläinen, A.-M.; Peet, A.; Tillmann, V.; Pöhö, P.; Mattila, I.; Lähdesmäki, H.; Franzosa, E. A.; Vaarala, O.; de Goffau, M.; Harmsen, H.; Ilonen, J.; Virtanen, S. M.; Clish, C. B.; Orešič, M.; Huttenhower, C.; Knip, M.; Xavier, R. J. The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes. *Cell Host Microbe* **2015**, *17* (2), 260–273. https://doi.org/10.1016/j.chom.2015.01.001.
41. Petroff, O. A. C. Book Review: GABA and Glutamate in the Human Brain. *The Neuroscientist* **2002**, *8* (6), 562–573. https://doi.org/10.1177/1073858402238515.
42. Li, S.; Li, Y.; Xue, C.; Zhang, Y.; Tong, T.; Ouyang, Z.; Liu, D.; Cai, J.; Sun, H. Progress in Research on the Mechanism of GABA in Improving Sleep. *Foods* **2025**, *14* (22), 3856. https://doi.org/10.3390/foods14223856.
43. Strandwitz, P. Neurotransmitter Modulation by the Gut Microbiota. *Brain Res.* **2018**, *1693*, 128–133. https://doi.org/10.1016/j.brainres.2018.03.015.
44. Aggarwal, S.; Ahuja, V.; Paul, J. Dysregulation of GABAergic Signaling Contributes to the Pathogenesis of Diarrhea-Predominant Irritable Bowel Syndrome. *J. Neurogastroenterol. Motil.* **2018**, *24* (3), 422–430. https://doi.org/10.5056/jnm17100.
45. Cruz-Sanabria, F.; Carmassi, C.; Bruno, S.; Bazzani, A.; Carli, M.; Scarselli, M.; Faraguna, U. Melatonin as a Chronobiotic with Sleep-Promoting Properties. *Curr. Neuropharmacol.* **2023**, *21* (4), 951–987. https://doi.org/10.2174/1570159X20666220217152617.
46. Hajak, G.; Rodenbeck, A.; Staedt, J.; Bandelow, B.; Huether, G.; Rüther, E. Nocturnal Plasma Melatonin Levels in Patients Suffering from Chronic Primary Insomnia. *J. Pineal Res*. **1995**, *19* (3), 116–122. https://doi.org/10.1111/j.1600-079x.1995.tb00179.x.
47. Li, Y.; Hao, Y.; Fan, F.; Zhang, B. The Role of Microbiome in Insomnia, Circadian Disturbance, and Depression. *Front. Psychiatry* **2018**, *9*, 669. https://doi.org/10.3389/fpsyt.2018.00669.
48. Li, W.; Wang, Z.; Cao, J.; Dong, Y.; Chen, Y. Melatonin Improves the Homeostasis of Mice Gut Microbiota Rhythm Caused by Sleep Restriction. *Microbes Infect.* **2023**, *25* (6), 105121. https://doi.org/10.1016/j.micinf.2023.105121.
49. Zhang, M.; Zhou, C.; Li, X.; Li, H.; Han, Q.; Chen, Z.; Tang, W.; Yin, J. Interactions between Gut Microbiota, Host Circadian Rhythms, and Metabolic Diseases. *Adv. Nutr.* **2025**, *16* (6), 100416. https://doi.org/10.1016/j.advnut.2025.100416.

50. Sutanto, C. N.; Loh, W. W.; Kim, J. E. The Impact of Tryptophan Supplementation on Sleep Quality: A Systematic Review, Meta-Analysis, and Meta-Regression. *Nutr. Rev.* **2022**, *80* (2), 306–316. https://doi.org/10.1093/nutrit/nuab027.
51. Lee, J.-H.; Hwang, S. J.; Ham, S. L.; Kim, J.; Bang, H. J.; Park, J.-S.; Jang, H.-H.; Kim, T. Y.; Park, J. W.; Seo, Y. R.; Kim, B. S.; Kim, G. S.; Lee, H.-J.; Kim, C. S. Gut Bacterial Metabolites from Tryptophan and Phenylalanine Induce Melatonin Synthesis and Extend Sleep Duration in Mice. *ACS Omega* **2024**, *9* (43), 43875–43883. https://doi.org/10.1021/acsomega.4c06923.
52. Hill, C.; Guarner, F.; Reid, G.; Gibson, G. R.; Merenstein, D. J.; Pot, B.; Morelli, L.; Canani, R. B.; Flint, H. J.; Salminen, S.; Calder, P. C.; Sanders, M. E. Expert Consensus Document. The International Scientific Association for Probiotics and Prebiotics Consensus Statement on the Scope and Appropriate Use of the Term Probiotic. *Nat. Rev. Gastroenterol. Hepatol*. **2014**, *11* (8), 506–514. https://doi.org/10.1038/nrgastro.2014.66.
53. Shah, A. B.; Baiseitova, A.; Zahoor, M.; Ahmad, I.; Ikram, M.; Bakhsh, A.; Shah, M. A.; Ali, I.; Idress, M.; Ullah, R.; Nasr, F. A.; Al-Zharani, M. Probiotic Significance of Lactobacillus Strains: A Comprehensive Review on Health Impacts, Research Gaps, and Future Prospects. *Gut Microbes* **2024**, *16* (1), 2431643. https://doi.org/10.1080/19490976.2024.2431643.
54. Bravo, J. A.; Forsythe, P.; Chew, M. V.; Escaravage, E.; Savignac, H. M.; Dinan, T. G.; Bienenstock, J.; Cryan, J. F. Ingestion of *Lactobacillus* Strain Regulates Emotional Behavior and Central GABA Receptor Expression in a Mouse via the Vagus Nerve. *Proc. Natl. Acad. Sci.* **2011**, *108* (38), 16050–16055. https://doi.org/10.1073/pnas.1102999108.
55. Yu, L.; Han, X.; Cen, S.; Duan, H.; Feng, S.; Xue, Y.; Tian, F.; Zhao, J.; Zhang, H.; Zhai, Q.; Chen, W. Beneficial Effect of GABA-Rich Fermented Milk on Insomnia Involving Regulation of Gut Microbiota. *Microbiol. Res*. **2020**, *233*, 126409. https://doi.org/10.1016/j.micres.2020.126409.
56. Zhao, N.; Shu, Y.; Jian, C.; Zhou, Z.; Bao, H.; Li, X.; Cheng, X.; Zhao, Y.; Jin, S.; Shu, X. *Lactobacillus* Ameliorates SD-Induced Stress Responses and Gut Dysbiosis by Increasing the Absorption of Gut-Derived GABA in Rhesus Monkeys. *Front. Immunol.* **2022**, *13*, 915393. https://doi.org/10.3389/fimmu.2022.915393.
57. Withrow, D.; Sprecher, K.; Gonzalez, A.; Livingston, F.; Noteboom, E.; Burke, T.; Depner, C. M.; Fleshner, M.; Lowry, C.; Turek, F.; Vitaterna, M.; Dorrestein, P.; Knight, R.; Wright, K. 0412 Better Objective Sleep in Healthy Adults Is Marked by Increased Bifidobacterium Breve and Saguini. *SLEEP* **2025**, *48* (Supplement_1), A180–A180. https://doi.org/10.1093/sleep/zsaf090.0412.
58. Moloney, G. M.; Long-Smith, C. M.; Murphy, A.; Dorland, D.; Hojabri, S. F.; Ramirez, L. O.; Marin, D. C.; Bastiaanssen, T. F. S.; Cusack, A.-M.; Berding, K.; Fouhy, F.; Allen, A. P.; Stanton, C.; Clarke, G.; Dinan, T. G.; Cryan, J. F. Improvements in Sleep Indices during Exam Stress Due to Consumption of a Bifidobacterium Longum. *Brain Behav. Immun. - Health* **2021**, *10*, 100174. https://doi.org/10.1016/j.bbih.2020.100174.
59. Allen, A. P.; Hutch, W.; Borre, Y. E.; Kennedy, P. J.; Temko, A.; Boylan, G.; Murphy, E.; Cryan, J. F.; Dinan, T. G.; Clarke, G. Bifidobacterium Longum 1714 as a Translational Psychobiotic: Modulation of Stress, Electrophysiology and Neurocognition in Healthy Volunteers. *Transl. Psychiatry* **2016**, *6* (11), e939–e939. https://doi.org/10.1038/tp.2016.191.
60. Patterson, E.; Tan, H. T. T.; Groeger, D.; Andrews, M.; Buckley, M.; Murphy, E. F.; Groeger, J. A. Bifidobacterium Longum 1714 Improves Sleep Quality and Aspects of Well-Being in Healthy Adults: A Randomized, Double-Blind, Placebo-Controlled Clinical Trial. *Sci. Rep*. **2024**, *14* (1), 3725. https://doi.org/10.1038/s41598-024-53810-w.
61. Zisapel, N. New Perspectives on the Role of Melatonin in Human Sleep, Circadian Rhythms and Their Regulation. *Br. J. Pharmacol.* **2018**, *175* (16), 3190–3199. https://doi.org/10.1111/bph.14116.
62. Adhish, M.; Manjubala, I. Effectiveness of Zebrafish Models in Understanding Human Diseases—A Review of Models. *Heliyon* **2023**, *9* (3), e14557. https://doi.org/10.1016/j.heliyon.2023.e14557.
63. Li, D.; Yu, S.; Long, Y.; Shi, A.; Deng, J.; Ma, Y.; Wen, J.; Li, X.; Liu, S.; Zhang, Y.; Wan, J.; Li, N.; Ao, R. Tryptophan Metabolism: Mechanism-Oriented Therapy for Neurological and Psychiatric Disorders. *Front. Immunol.* **2022**, *13*, 985378. https://doi.org/10.3389/fimmu.2022.985378.
64. Tsuji, A.; Ikeda, Y.; Yoshikawa, S.; Taniguchi, K.; Sawamura, H.; Morikawa, S.; Nakashima, M.; Asai, T.; Matsuda, S. The Tryptophan and Kynurenine Pathway Involved in the Development of Immune-Related Diseases. *Int. J. Mol. Sci*. **2023**, *24* (6), 5742. https://doi.org/10.3390/ijms24065742.
65. Vaghef-Mehrabany, E.; Maleki, V.; Behrooz, M.; Ranjbar, F.; Ebrahimi-Mameghani, M. Can Psychobiotics "Mood" Ify Gut? An Update Systematic Review of Randomized Controlled Trials in Healthy and Clinical Subjects, on Anti-Depressant Effects of Probiotics, Prebiotics, and Synbiotics. *Clin. Nutr.* **2020**, *39* (5), 1395–1410. https://doi.org/10.1016/j.clnu.2019.06.004.

Authors

Poorvi Kurundkar, currently studying in the International Baccalaureate (IB) program at Oakridge International School, is deeply fascinated by how the brain orchestrates complex physiological systems. As a result, she aspires to pursue a career in neuroscience-based research.

DOI: 10.36838/IJHSR819.13

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Is AI's Threat to Existential Unemployment Technologically Deterministic or Contingent on AI Governance?

Rayan Moodley

St. John's College, St David Rd, Houghton Estate, Johannesburg, Gauteng, 2198, South Africa; rayanmoodley001@gmail.com

ABSTRACT: Since the dawn of the rapid expansion of generative Artificial Intelligence (AI) technologies in 2022, companies have started using generative machine learning (ML) AI systems in the workplace for various tasks. Due to the surge in AI usage, many people have grown skeptical of AI, viewing it as a threat to unemployment and human contribution in the workplace. This creates a dilemma where AI is limited in its capabilities to assist people in the workforce, as it is widely viewed as a form of human replacement. In this article, I will explore the extent to which AI constitutes an existential threat to employment and whether AI can be structured to augment and assist rather than replace human workers. I will conduct a secondary analysis of post-2022 labor-market data. Some studies argue that while AI exacerbates existential unemployment through task substitution and labor displacement, its long-term impact is contingent upon policy creation aiming at employing ethical AI in the workplace and effective human–AI collaboration structures. By identifying gaps in current research and integrating ethical and existential perspectives with past labor metrics, this study reframes AI not as an inevitable replacement for human labor but as a technology whose employment effects remain fundamentally malleable and contingent upon its ethical application.

KEYWORDS: Intelligent Machines, Cognitive Systems, Ethical AI Governance.

Introduction

The generative AI expansion that dawned in late 2022, such as ChatGPT artificial intelligence (AI), has transitioned from a niche computational system to a force of change in the world. Firms have accelerated their development of machine learning (ML) Large Language Models (LLM's) capable of performing jobs once thought to be exclusive to human workers.[1] These AI systems are now used for tasks such as writing, coding, data analysis, and more. While this creates optimism about productivity, AI simultaneously exists as a growing concern as a potential threat to existential unemployment and the erosion of human contribution in the workforce.

This paper examines the extent to which generative AI exists as an existential threat to employment and whether the use of ethical AI and institutional AI policies can promote human-AI collaboration instead of human replacement. This paper will use post-2022 economic and labor-market studies, alongside AI ethics and policy literature, to assess AI's impact on unemployment.

Some empirical studies indicate AI is altering how the workforce is structured, either by augmenting or replacing humans. Eloundou *et al.* estimate that approximately 80% of U.S. workers could see at least 10% of their tasks affected by large language models, with around 19% potentially experiencing impacts on more than half of their job, particularly in text-based and data-analytical roles.[2] Global patterns mirror this trend, as evidence from the International Labor Organization shows that a substantial part of global employment is at risk of being affected by generative AI at varied levels, emphasizing AI's influence on job transformation rather than job elimination.[3,4] These findings show that evaluating AI's effect on labor markets based on aggregate employment metrics is not an accurate representation of AI's influence and requires the use of task-level metrics to assess its influence.

Three quantitative metrics will be used throughout this paper to accurately assess this. The metrics include the Task Automation Exposure Indices developed by Eloundou *et al.*,[2] which measure the proportion of job duties theoretically susceptible to AI augmentation. These rely on the structure of tasks within specific occupations. Skill Obsolescence rate, which is grounded in longitudinal labor research like that done by Deming and Noray,[5] measures the rate at which once-valuable human skills decline in relevance as AI capabilities expand. Labor Market Polarization Metrics, which build on work done by Autor, Levy, and Murnane,[6] reveal how employment and wage distributions are shifting across skill and occupation categories in response to technological change, suggesting that high-skill and low-skill jobs are affected differently. This pattern shows uneven effects across occupations. As automation advances, income gaps may widen alongside employment trends. Though not all regions experience this equally, the overall distribution reflects structural change.

These three metrics will be used as an analytical framework rather than independent datasets. Task exposure indices will be used to evaluate which job categories are most vulnerable to AI substitution, while skill obsolescence rates will be used to assess how quickly human competencies lose relevance over time. Finally, Labor market polarization metrics are to be used to interpret wage distribution and employment shifts across skill groups. Together, they allow cross-comparison of structural labor-market change rather than simple employment-level effects. This creates a more detailed analysis of AI's effect.

Recent research papers support this structural change by using these metrics, which they position as fundamental in de-

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.14

termining AI influence. Originally done by Eloundou *et al.*, then improved upon by the ILO, the task automation exposure index combines task-specific data to measure AI's influence across different jobs and countries.[4] The skill obsolescence rate metric: Over time, shifting demands for certain abilities are captured using extended workforce studies revealing AI's influence on human-based competencies. The Labor polarization metric features both employment analysis and wage distribution tests to measure structural changes in job standards and pay gaps that emerge through methods like those applied by Ganuthula and Balaraman,[7] who examine hiring patterns alongside income spread between high-skill and low-skilled workers. When combined, such metrics reveal how AI may be having a multifaceted effect on transformations in job structures, with some positions becoming less important and others adapting or emerging.[4,7]

However, analyses including reports by the OECD with broader labor demand data indicate that human dependency has not dropped sharply due to artificial intelligence.[8] Instead, variations start to emerge when examining specific industries and professions, where impacts cluster around distinct worker categories and narrowly defined automated activities. In contrast, early studies based on digital trace data and real-world experiments reveal improved output in select work environments while also detecting shifts replacing people with algorithms during elementary composition and linguistic conversion duties.[9] These complexities and frankly contrasting findings clarify that it is insufficient to view AI's impact based solely on employment counts. Over the course of this paper, I will prove how detailed task-level metrics and exposure indices are essential to understanding the nuances of AI's influence.

This paper asserts that AI can exacerbate existential unemployment through task substitution, but its long-term employment effects are contingent upon institutional design, ethical AI application, and AI-policy frameworks, such as the EU AI Act and the NIST AI Risk Management Framework. The Theoretical Framing section will outline the foundations of this argument, followed by a dedicated methodology section which will provide insight into the paper's structure. In choosing to anchor this analysis in labor metrics and ethical AI policymaking, we aim to clarify whether AI's long-term impact is a technologically deterministic threat to human employment or contingent upon ethical AI application.

Theoretical Framing

The debate surrounding AI and employment is not purely empirical but also theoretical. An interesting and influential perspective comes from Penrose, who argues that human cognition involves non-computable processes that cannot be completely replicated by machines.[10] His claim suggests that logical reasoning, specifically intuition, consciousness, and deep understanding, may remain resistant to automation.

In contrast, task-based economic models of automation conceptualize AI as a system capable of substituting specific tasks rather than entire jobs. Acemoglu and Restrepo argue that technological change restructures labor demand by reallocating tasks between humans and machines rather than eliminating work.[11] These frameworks align with empirical findings showing that AI redistributes labor rather than fully replacing it.

This paper adopts a hybrid perspective: while AI may replicate structured cognitive tasks, there remain specific aspects of human labor, such as judgment, adaptability, and ethical reasoning, that are less susceptible to automation. This theoretical framing supports this paper's central claim that AI's impact on employment is not technologically deterministic but contingent on how technological capabilities interact with human and institutional factors.

Methodology

In this paper, I will analyze and review whether this statement is true. In Section 1, we will conduct a secondary analysis on AI's negative impact on employment, with Section 1.1 examining task automation exposure, Section 1.2 analyzing skill obsolescence trends, and Section 1.3 assessing labor-market polarization effects. Section 2 evaluates how ethical AI policies and human-dependent work policies influence employment outcomes, while the Discussion will outline how this research helps policymakers, educators, employers, and technology developers make decisions that could shape future labor markets and AI ethics to further the development of AI. Furthermore, the findings will help inform workers, especially in the younger generation, of AI's impact on whether employment remains a consistent form of income, social stability, and human dependency or whether alternative models of human contribution must be developed.

Previous studies around automation concentrated on replacing repetitive tasks and repolarizing the workforce due to computerization and industrial automation. A number of these studies, however, have been done before the advent of generative artificial intelligence (AI) systems that are now being developed to execute cognitive or language-based functions previously thought to be unachievable automatically. In the same way, when measuring skill obsolescence, some current research does not account for the role of AI exposure in the task, nor for the impact of an increasing number of agentic AI systems with some level of autonomy. In turn, the previous models and frameworks might underestimate how much generative AI will impact professional and cognitive work. In this paper, we aim to overcome these limitations by integrating the notion of polarization, skill obsolescence, and Task Automation Exposure measures in a single framework. As a result, the methodology allows a more comprehensive assessment of the impact of AI on the traditional and thinking man's jobs.

The three selected metrics were chosen because together they capture different dimensions of AI-driven labor transformation with respect to overlooked yet crucial statistics. Labor market polarization metrics explain long-term structural shifts in wages and employment distribution,[6] the skill obsolescence rates measure how rapidly worker competencies lose relevance in changing industries.[5] and the Task Automation Exposure metrics were selected because they directly evaluate how generative AI affects individual job tasks rather than entire occupations. This is particularly important as it highlights

which specific aspects of the workspace AI influences more than others.[2] Past research was targeted on the impact of AI on the labor market in one area only, which is restrictive for a more comprehensive study. However, this paper brings together the three metrics for comparing the vulnerability of tasks, workforce adaptability, and restructuring of employment. The analysis is broken down by Levels of skill, Automation risk, and Human Oversight of the work performed. High-skilled jobs include those involving in-depth critical thinking, like a software analyst, and medium-skilled jobs include people in administration or working in a call center: low-skill jobs include workers on a construction site or factory. While every metric focuses on a different dimension of the changing course of the job market, the combination helps provide a larger context for understanding the impact AI is having on existential unemployment. Labor polarization indicators provide an insight into the structure of employment and wage distribution, and skill obsolescence measures uncover the dynamics of the loss of vital skills due to technological progress. Additionally, Task Automation Exposure metrics can be used in conjunction with either approach to help determine the tasks within an occupation that are most vulnerable to augmentation or automation. There are, however, conflicting signals among these indicators: the level of employment appears to be unchanged, while disruptions and restructuring among the tasks in which workers engage increase. There are lots of interesting comments. There are some interesting comments here. All three methods are integrated into the paper, which provides a more comprehensive evaluation of AI than studies focused on a single measure, by considering it at three different levels (the task, the skill, and the job market). Therefore, this three-layer structure will support the conclusions that were made in the paper and will lead to a stronger conclusion.

Section 1: AI's Impact on the Job Market

Overview:

The current job market is faced with a potential threat from task-level automation and how its implementation in the workforce causes people to be seen as having skills that are easily replaceable. As a result, this forces a change in labor markets. Thus, we see transformations in job structures with some positions fading and others adapting or emerging. Whether this change is positive or negative is up for debate, but the consensus thus far is that AI is undoubtedly sparking a change in global job structures, potentially being a catalyst for the fourth industrial revolution.[9]

Section 1.1.: How Much AI Can Automate Tasks:

Job Task automation has become more viable and applicable in recent years due to the rapid advances in generative AI. Research shows that automation effects are limited to specific job tasks, not entire jobs, and that differences between the two are seen when comparing the differences between high-skill, mid-skill, and low-skill jobs.

Task Automation Exposure provides quantitative evidence effectively proving that AI affects occupations differently depending on task structure. According to Eloundou *et al.*, high-skill jobs such as software developers and financial analysts show exposure rates above 40% because many analytical and coding tasks can be augmented by generative AI. Mid-skill occupations, such as administrative assistants, experience similarly high exposure because repetitive documentation and scheduling tasks can be automated. These classifications are supported by the OECD's occupational frameworks, which categorize occupations according to educational requirements, task complexity, and adaptability.[8] This demonstrates that AI vulnerability depends less on job titles themselves and more on whether the occupation relies on repetitive cognitive tasks.

LLM's demonstrate significant capacity with job-related task records. Recent scholarship results indicate 80% of American jobs include at least 10% of duties exposed to automated handling; furthermore, 19% face over half their jobs within reach of such systems. This shows that tasks exposed to potential automation are centered on segmented activities, examples such as AI writing emails or documents, condensing and analyzing data, or assisting with programming.[2] This highlights that AI affects specific tasks within a profession but not the whole profession. Due to this pattern, jobs involving menial tasks, data analysis, or any form of procedural tasks show higher susceptibility to replacement. Thus, generative AI appears better suited to alter job structures rather than replace entire positions.

Global evidence backs a focus on task-level automation when assessing how AI supports human work. According to the International Labor Organization, job transformation and the shift of menial tasks to be automated happen more frequently than outright job replacement, especially in administrative positions involving repetitive cognitive analysis.[3,4] A detailed worldwide measure of occupational exposure shows how most jobs combine both automatable tasks and tasks that are human-dependent, effectively creating an ecosystem of human-AI collaboration.

However, research now shows an opposite trend where jobs that once required advanced skills may be able to resist general replacement but face a growing risk of partial augmentation by AI. The 2023 OECD Employment Outlook highlights how progress in understanding human language has enabled LLM's to start to oversee complex thinking tasks once exclusive to humans. Instead of AI fully replacing humans in their professions, certain duties or tasks shift to being task-automated, especially in fields like finance, law, management, and engineering. The work that humans do does not vanish, but rather it moves toward human-dependent competencies such as integration and judgment. Tasks end up changing rather than disappearing or being completely automated by AI. This reinforces the claim that AI functions as a task-changing technology rather than a mechanism for complete human replacement.

Additionally, labor market evidence reflects the reality that job tasks end up changing rather than being replaced. From an analysis of millions of digital job postings by Mäkelä and Stephany,[12] they discovered evidence not only of automatable tasks but also a rising need for advanced human skills tied to AI. The study shows that tasks such as troubleshooting, data collection, and organizing gain importance as efficiency improves

DOI: 10.36838/IJHSR819.14

across sectors. This effect appears strongest in roles requiring advanced expertise, where technology serves alongside workers rather than substituting them, especially regarding generative systems. This demonstrates that AI creates a parallel demand for complementary human skills, supporting the augmentation hypothesis rather than full automation.

Low-skilled jobs face a much greater risk of being replaced by AI. A study by Zhang *et al.* shows that when companies introduce AI into low-skill work environments, they often restructure tasks into rigid, standardized procedures and subject them to potential automation.[13] As a result, worker autonomy declines even when jobs are not fully automated. Generative AI affects labor across all skill levels, but displacement risk falls unevenly. It strongly concentrates among low-skilled workers, while higher-skilled workers typically adjust by integrating AI into their workflows rather than being eliminated.

Section 1.2.: Skill Obsolescence Rates:

Continuing from the analysis of task-level automation impacts mentioned in Section 1.1, another way that generative AI contributes to shifts in the workforce is due to the fading relevance of specified abilities, particularly those involving structured thinking or consistent logical processes. Rather than eliminating positions, AI alters what skills are valuable across roles. Some rapidly lose importance, whereas others can gain prominence and perceived value. Dynamics such as this can play a central role in reshaping how occupations evolve by continuously redefining what skills retain or lose economic value.

From long-term workforce research, signs of change emerge. As shown by OECD data from 2023, when jobs involve AI more frequently, the needed skills shift at a similar pace, especially where standard analytical or administrative tasks begin fading. Instead of enduring value, competencies now fade sooner once tools take over. Skills lose relevance faster under AI integration, reshaping what workers must know well before mastery becomes outdated.

This perspective is backed by established economic research. Workers who specialize in fast-changing technological areas face faster skill depreciation, according to Deming and Noray,5 which leads to lower earnings if knowledge remains static. Although their analysis came before the surge of AI, it holds relevance today because such technologies accelerate obsolescence by overseeing narrow objectives and tasks. As a result, the ability to learn and adjust gains greater value. In parallel, Acemoglu and Restrepo11 have observed that when AI takes over certain job functions, employment shifts toward different activities – ones requiring new competencies. Without movement into these emerging positions, expertise loses usefulness and risks becoming redundant.

Evidence that ties closely to AI, reinforcing earlier conclusions from Mäkelä and Stephany,12 shows a rising need in employment for abilities that go alongside AI, such as solving complex issues or navigating digital tools, while requirements for routine, rule-based competencies shrink. What stands out is how outdated skills emerge unevenly within economies shaped by automation, falling hardest on those whose expertise mirrors what machines now oversee.

When viewed collectively, these studies show how AI enables one set of abilities to replace others, leading to faster changes in such competencies instead of simple job loss. Such patterns reinforce the idea that introducing AI into work environments acts as a trigger for deeper shifts across job structures, thus prompting a need for adaptation in human capabilities and task assignments to better suit human competencies that AI cannot replicate. This shift forms the idea of whether AI has the potential to be a catalyst for a fourth industrial revolution.

Section 1.3.: Labor market polarization:

Labor market polarization emerges when automation at the task level coincides with faster erosion of mid-level skills. Employment and wage growth shift toward both high-skill and low-skill positions, whereas roles requiring moderate expertise and capabilities shrink. Shifts such as this reveal how generative AI alters work dynamics and thus potentially boosts output and efficiency in some sectors, which in turn widens disparities in other sectors. The outcome is fundamentally a structural change marked by uneven rewards across different skill bands.

Research from Murnane *et al.* explains the cause for job market splits due to technology taking over menial duties, while also supporting roles that need critical thinking skills or physical effort, which are fundamentally human-dependent.[6] Where machines once altered repetitive work, generative artificial intelligence has reached the ability to perform cognitive tasks once thought safe from automation. Section 1.1 highlights this shift, where intelligent systems can now reshape complex thinking roles and thus increase strain on mid-level professional positions that rely heavily on performing such tasks.

The expansion of AI into cognitive tasks changes the original understanding of labor polarization because automation is no longer limited to repetitive physical work. Generative AI systems can now perform legal drafting, medical documentation, and data interpretation tasks once considered resistant to automation. However, these systems still face major reliability limitations. For example, lawyers using ChatGPT in legal research have submitted hallucinated court cases generated by AI systems, creating serious professional and ethical risks.[9] Similarly, autonomous vehicle systems continue to require human supervision because algorithmic errors can contribute to fatal accidents.[9] These examples show that many AI systems remain semi-autonomous rather than fully independent. As a result, human oversight continues to be necessary even when AI performs advanced cognitive functions.

A study by Ganuthula and Balaraman,[7] which analyses job trends alongside wage distribution, backs this change, as the data reveals a link between AI integration and growing disparities in earnings across different skill groups. With workforce statistics and measurements of wage differences, the results show that those with advanced skills benefit more from efficiency boosts tied to AI, whereas positions requiring moderate expertise see limited growth or even shrinking opportunities. Such findings align closely with the framework on employment distribution presented here. This indicates that AI does

not affect all workers in the same way. It benefits high-skill labor while compromising opportunities in mid-skill occupations.

Findings gain support from institutional research. According to the ILO,[4] clerical and administrative positions typically held by mid-level skilled workers are more heavily impacted by AI, whereas higher-skilled roles see changes in duties instead of job loss. On the opposite end, lower-skilled service work continues, though pay advances are rare, contributing to a divided job market. In parallel, the OECD[8] observes wider gaps in earnings following AI integration, despite a minor change in total hiring numbers.

Few roles shift as distinctly as those requiring advanced abilities and skills. These positions do not vanish but instead transform into a system where AI supports output and overall efficiency while promoting and assisting human-dependent jobs that value AI insight, decision-making, and long-term planning. Meanwhile, routine-based occupations encounter a greater risk of replacement or reduced responsibilities, leading to a perception of lower uniqueness and an increase in potential replaceability in required competencies. Such differences reveal an outcome where gains in performance coexist with widening job disparities.

Together, this evidence suggests that shifts in job demand and the fading relevance of certain skills show how AI affects current work structures. Although AI may not directly cause widespread job loss, it can alter how people find employment, change the nature of how income is distributed, and speed up large-scale transitions within the workforce, which highlights its role as a major influence instead of just another technological tool. All the data points to AI having different effects on job markets based on the level of competency, i.e., high-skill, middle-skill, or low-skill jobs.

Section 2.: Governance, Ethics, and Job Outcomes:

Overview:

The decisive aspect of ethical AI governance is the way in which generative artificial intelligence will lead to existential unemployment or help in changing jobs more adaptively. Instead of thinking about labor displacement as an inevitable technology, the current policy structures and critical literature highlight that the interaction between AI systems and human workers depends on institutional decision-making. The governance mechanisms determine either the replacement of labor by AI, the reorganization of duties, or the augmentation of human input in companies.

Section 2.1. EU AI Act:

The EU AI Act is an example of a policy created to combat the threat of AI in the workforce. It is a regulatory framework that regulates high-risk AI systems regarding aspects of employment, recruitment, and worker management to ensure safety, transparency, and human oversight. Jobs that involve any form of recruitment, worker appraisal, worker task assignment, and performance surveillance by AI tools fall under a high-risk category of AI replacement. This triggers obligatory demands where transparency, human control, and accountability are required from companies that utilize AI. These measures restrict companies from employing and utilizing entirely autonomous decision-making approaches in the management of labor and thus can virtually guarantee that humans have final control over their employment consequences.[14] This reduces the likelihood of full automation in employment decisions and reinforces human accountability in workplace systems.

The fact that it entails the protection of human intervention and an ability to explicitly document the behavior of the system lowers the chance that workers will be seen as redundant in digitalized workplaces such as software engineering. This type of governance changes AI applications to be more of a contribution to jobs rather than a mechanism for replacing human workers, which in turn supports the previous research that AI ends up changing job structures more frequently than it takes over jobs.[2,4] Through this, we can see that regulation serves as a measure to protect against existential unemployment because it maintains worker agency and occupational identity. Furthermore, the evidence proves that human oversight strengthens the protection of workers but simultaneously introduces compliance trade-offs. Therefore, employment outcomes depend on the efficiency with which AI regulations can balance both AI innovation and human worker safeguarding.

Section 2.2. NIST AI Risk Management Framework:

The NIST AI Risk Management Framework is another regulatory solution to AI's threat, as it offers a guided framework to companies on the implementation of AI responsibly. It points to the displacement of the workforce, deskilling, and pivoting away from human control as risks that cannot be tolerated but must be actively addressed. The framework focuses on accountability frameworks, situational implementation, and ongoing examination of the impacts of AI on human value and contribution in the workplace.[15]

This strategy is supported by empirical studies of labor, indicating that productivity increases are greater when AI supplements skilled labor instead of substituting it. In situations where institutions develop AI systems that assist human judgment, coordination, and long-term reasoning, human workers continue to play a leading role in companies. On the other hand, inefficient governance will put automation at risk of contributing to the decline of the quality of jobs and hastening polarization.[8,12] Thus, the argument that employment results are due to governance decisions and not technological capability alone is strengthened by this framework.

Section 2.3.: AI Policy Development and Its Implications in the Workplace:

The creation of AI policies is at the center of the argument on whether AI will supplement or cause existential unemployment among those employed by companies utilizing AI. Researchers have claimed that the governing structures, regulatory frameworks, and discourse surrounding AI determine the way organizations apply AI in the workplace.[16,17] Policies that emphasize human supervision, ethical implementation, and workplace transparency with AI minimize the threat of existential unemployment. This shows how Ethical AI policies

DOI: 10.36838/IJHSR819.14

create a framework that utilizes AI to its fullest potential while mitigating it as a threat.

Furthermore, studies show that companies viewing AI's threat as deterministic rather than dependent on policies are at risk of increasing the polarization of labor and skill obsolescence.[16] Conversely, companies that adhere to ethical AI models place greater emphasis on human-AI partnership, distribution of tasks, and skills acquisition, which makes the workers complement AI instead of being replaced by it.

The discourse is also important to note, as the ways that policymakers, the media, and corporate executives define AI will influence the degree to which the population and organizations perceive AI as a threat or a productivity tool.[17]

These findings support my paper's claim that the effects of AI on the employment sector are not deterministic but policy contingent: outcomes in the workplace are influenced by regulatory frameworks, ethical principles, and social dialogue, which demonstrates that existential unemployment is not so much a consequence of AI technology but rather a product of policy and governance decisions.[16]

Discussion

Based on prior evidence analyzed in this paper, it can be argued that AI is transforming workplace frameworks by reallocating tasks, not by replacing humans and causing mass unemployment. As AI systems take the responsibility of basic cognitive tasks, human workers are playing a more prominent role in diplomatic thinking, synthesizing, and strategic decision-making, increasing productivity and the efficiency of their companies and in their respective fields. These changes help in economic expansion and a change in the definition of what constitutes valuable labor.[8]

However, it is important to acknowledge the limitations of this paper. Firstly, as a review article, it is reliant on pre-existing datasets and interpretations with differs across sources. Moreover, a large portion of the data and analyses used in this paper were conducted in developed economies, which hampers the generalization of the results across the globe, particularly in how technological impact is different in developing countries. Lastly, the rapidly evolving nature of AI also means that current trends and research may not fully capture AI's future capabilities, as well as how the labor market responds.

Nonetheless, the difficulties of adapting to AI are a potentially serious threat. Employees who cannot learn new skills that are AI-complementary become less relevant, can suffer increased task control, and are less likely to feel irreplaceable, especially with low-skill jobs where automation consolidates authority over employees.[13] This effect is uneven, with low-skill workers facing higher displacement risk compared to high-skill workers who benefit from AI augmentation. Meanwhile, the wage disparity expands as productivity increases are obtained by employees and companies that are in the best position to adapt AI into their workforce successfully.[7] Additionally, across the three metrics of task exposure, skill obsolescence rate, and the data regarding wage distribution, the evidence suggests that AI is more inclined to restructure tasks in the workplace rather than independently influence employment outcomes. It is important to note that AI will also create new job opportunities, and unemployed people will be able to relocate themselves accordingly. Unfortunately, lower-income individuals or individuals with lower education, historically, are slower to adapt to modern technologies or shifts in how the job market is structured.[18]

These findings overall underscore the importance of human oversight in the workforce; they emphasize the value of flexibility within the workforce as technology is constantly evolving, and the only way to ensure a company's survival is to adapt. For employers and developers, these findings emphasize the importance of creating AI systems that complement human contribution rather than replace it. Finally, the evidence presented throughout this paper proves that the question of whether AI will create existential unemployment is not as much about the state of technological development as the question of whether the ethical decisions of governance will dictate how work, human societal value, and human purpose will change in an AI-driven economy.

Conclusion

Throughout this paper, evidence and arguments have been presented to determine whether AI as an existential threat to unemployment was dependent on its ethical application rather than its technological advancement. The data and policies used from other research papers show that AI does not fully eliminate human workers but rather reallocates which tasks are done by humans and which ones are AI-dependent. It has also been proven that AI has accelerated skill obsolescence in routine tasks and reshaped job structures unevenly across skill gaps.

It has been proven that at a task-level, exposure analyses indicate that AI is mostly used to automate specific cognitive tasks instead of entire occupations, showing that AI applications currently result in job transformation rather than job loss. However, it has been proven that many low-skill workers not only face increased AI control but also risk of total replacement due to AI making their competencies obsolete they can be automated, which can lead to the polarization of the labor market. Furthermore, it was shown that there is a high likelihood of high-skill workers benefiting due to productivity improvements brought about by AI, as well as the erosion of mid-skill jobs. These trends tend to be obscured by aggregate employment statistics, which also highlights the earlier point that task-level and distribution metrics are key in understanding AI's influence.

The most crucial point made was that the review shows that such results are not predetermined by AI's advanced capabilities over humans, but rather its ethical implementation. The current government frameworks presented, such as the EU AI Act and the NIST AI Risk Management Framework, are just two examples that demonstrate the potential for AI policies as a catalyst to not only protect human workers but also show that AI need not be a tool for fear-mongering but instead an innovative technology with the potential for societal advancement.

It is ethical AI policies and institutional designs that therefore control future employment outcomes.

Overall, the effects of AI on the labor market can be viewed not as an inevitable catalyst for existential unemployment, but rather as a modern technology whose ethical governance is the deciding factor of whether AI's influence on the job market is assistive or an object of replacement. Future research should examine how AI-driven labor shifts differ across developing and developed economies, particularly in relation to inequality and access to reskilling opportunities.

Acknowledgments

I would like to acknowledge the following individuals who helped me write this paper. Their input and knowledge in the field of artificial intelligence were invaluable: Dr. Samuel Lefchort, Dr. Siddarth Krishnan (faculty member of Computer Science at UNC Charlotte with 10+ years in Artificial Intelligence and Machine Learning research), Dr. Audrey Wozniak, and Kieran Tait.

References

1. Wu, T.; Zhu, H.; Albayrak, M.; Axon, A.; Bertsch, A.; Deng, W.; *et al.* LLMs as Workers in Human-Computational Algorithms? Replicating Crowdsourcing Pipelines with Large Language Models. *arXiv* **2023**. https://arxiv.org/abs/2307.10168.
2. Eloundou, T.; Manning, S.; Mishkin, P.; Rock, D. GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. *Natl. Bur. Econ. Res. Working Paper* **2024**. https://arxiv.org/abs/2303.10130.
3. International Labor Organization. Generative AI Is More Likely to Augment Jobs than Destroy Them; ILO: Geneva, 2024. https://www.ilo.org/resource/news/generative-ai-likely-augment-rather-destroy-jobs.
4. International Labor Organization. Generative AI and Jobs: A Refined Global Index of Occupational Exposure, Working Paper No. 140; ILO: Geneva, 2025. https://www.ilo.org/publications/generative-ai-and-jobs-refined-global-index-occupational-exposure.
5. Deming, D. J.; Noray, K. Earnings Dynamics, Changing Job Skills, and STEM Careers. *Q. J. Econ.* **2020**, *135* (4), 1965–2005. https://doi.org/10.1093/qje/qjaa021.
6. Autor, D. H.; Levy, F.; Murnane, R. J. The Skill Content of Recent Technological Change. *Q. J. Econ.* **2003**, *118* (4), 1279–1333. https://doi.org/10.1162/003355303322552801.
7. Ganuthula, R.; Balaraman, V. AI Exposure, Hiring Patterns, and Wage Inequality. *J. Labor Econ.* **2025**. http://arxiv.org/abs/2501.15809.
8. Organization for Economic Co-operation and Development. *OECD Employment Outlook 2023: Artificial Intelligence and the Labor Market*; OECD Publishing: Paris, 2023. https://www.oecd.org/en/publications/oecd-employment-outlook-2023_08785bba-en.
9. Freund, L. B.; Mann, L. F. *Job Transformation, Specialization, and the Labor Market Effects of AI*. CESifo Working Paper; CESifo: Munich, Germany, 2025.
10. Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*; Oxford University Press: Oxford, 1989.
11. Acemoglu, D.; Restrepo, P. Robots and Jobs. *J. Econ. Perspect.* **2020**, *34* (3), 30–55. https://www.journals.uchicago.edu/doi/10.1086/705716.
12. Mäkelä, M.; Stephany, F. Complement or Substitute? How AI Increases the Demand for Human Skills. *arXiv* **2024**. https://arxiv.org/abs/2412.19754.
13. Zhang, Y.; Li, X.; Chen, H. Artificial Intelligence, Task Automation, and Labor Control: Evidence from Low-Skill Occupations. *J. Chin. Sociol.* **2025**, *12* (1). https://doi.org/10.1186/s40711-025-00249-9.
14. European Union. *EU Artificial Intelligence Act*; Official Journal of the European Union, 2024. https://artificialintelligenceact.eu/.
15. National Institute of Standards and Technology. *AI Risk Management Framework (AI RMF 1.0)*; U.S. Department of Commerce, 2023. https://www.nist.gov/itl/ai-risk-management-framework.
16. O'Brien, R. Play, and No Work? AI and Existential Unemployment. *J. Bus. Ethics* **2025**. https://doi.org/10.1007/s10892-025-09522-y.
17. Westerstrand, A.; *et al.* Talking Existential Risk into Being: A Habermasian Critical Discourse Perspective to AI Hype. *AI Soc.* **2024**. https://doi.org/10.1007/s43681-024-00464-z.
18. Sy, I. *The Future of Work: How Artificial Intelligence Is Shaping Job Markets and Employment Trends*. Preprints.org 2025, Version 1, May 17, 2025.
19. Qin, Y. *et al.* Artificial Intelligence: Intensifying or Mitigating Unemployment? *Technol. Soc.* **2024**. https://doi.org/10.1016/j.techsoc.2024.102755.
20. Ahmed, A. *Artificial Intelligence and the Skills Gap* — Kenan Institute Report; Kenan Institute of Private Enterprise, 2025.
21. Fruits, E.; Stout, K. AI, Productivity, and Labor Markets: A Review of the Empirical Evidence. *Int. Ctr. Law Econ.* **2026**, February 5, 2026. https://laweconcenter.org/resources/ai-productivity-and-labor-markets-a-review-of-the-empirical-evidence/.

Author

Rayan Moodley is a senior in high school. He enjoys learning about AI and technology, and found his journey in writing this paper to be a transformative experience. He hopes to attend UC Berkeley and plans to major in Engineering or AI.

DOI: 10.36838/IJHSR819.14

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

A Review of the Bilingual Advantage on the Onset of Dementia: The Knowledge Gap of Late L2 Acquisition

Rio Imawaka

Carlmont High School, 1400 Alameda de las Pulgas, Belmont, California, 94002, USA; rio.imawaka@gmail.com

ABSTRACT: There has been an ongoing controversy over the so-called "bilingual advantage," which is the advantage bilinguals have in executive functioning compared to monolinguals. One advantage in this field is the impact of bilingualism on delaying the onset of dementia. Many previous studies have investigated lifelong bilingualism and its advantages in this syndrome, but little research has examined bilinguals who acquired their second language (L2) later in life. Results from several studies that focus on this group are inconsistent due to differences in the definitions of bilingualism, making the data incomparable. Therefore, the purpose of this review is to address whether there is evidence of a bilingual advantage in delaying the onset of dementia for late bilinguals, and to identify potential improvements in research methods. Overall, better definitions of bilingualism, including age of language acquisition, proficiency, language differences, and daily language use, are needed. Once researchers investigate late bilingual advantage more and either prove or disprove it, we can get a better understanding of the bilingual advantage in cognitive diseases and make better suggestions for L2 learning in order to eventually decrease the number of dementia patients.

KEYWORDS: Behavioral and Social Sciences, Cognitive Psychology, Bilingualism, Executive Functions, Dementia, and Age of Acquisition.

Introduction

Multiple studies indicate that some factors, such as a healthy diet,[1] cognitively stimulating activities,[2] and high levels of education,[3] might prevent or delay the onset of dementia. Studies in the field have suggested that bilingualism might be one of them.[4–6] There is an ongoing controversy on the existence of the so-called bilingual advantage in non-linguistic domains.[5]

Bilingual advantage is the advantage in executive functioning for those with the ability to speak and use two languages.[7] Some studies report findings that support this theory,[5,8] while some refute or partially refute it.[7,9] For example, Bialystok *et al.* discovered through different tasks that lifelong bilingualism may have an advantage in executive functioning skills such as inhibition and working memory when compared to monolinguals.[8] Conversely, Paap and Greenberg claim that there is no evidence of bilingual advantage over monolinguals in executive control skills since the advantages were not consistent across different tasks that indicate executive functioning skills.[9] These inconsistencies across findings are the cause of the debate in the research field.

One of the non-linguistic factors that bilingualism might have a positive impact on is the onset of dementia.[5] It is believed that bilingualism can delay this,[6] but some limitations or gaps need to be better accounted for to make this claim more generalizable.[10]

One of the limitations is that existing research has mainly focused on lifelong bilinguals.[5,6] Due to this, the findings of these studies might only be generalizable to certain groups. One of the populations that needs more attention is the group of late bilinguals, or bilinguals who attained their second language (L2) later in life. Increased attention on late bilinguals in the research field will contribute to identifying what types of bilingualism provide the benefit of a delayed onset of dementia. This improvement will also contribute to the need for deeper consideration of individual differences, which is suggested by De Bruin.[10] If future findings support the bilingual advantage, then suggestions can be made on L2 education for young children and older adults. Also, if these suggestions contribute to a decrease in the number of dementia patients, then it would benefit the global economy, since dementia care costs 1.3 trillion USD in 2019, and the cost is predicted to increase as years pass.[11]

Ultimately, the purpose of this review is to summarize the findings on one of the many proposed ways of delaying dementia and to encourage accounting for different types of bilinguals in future research. In an effort to do so, this review first discusses and provides a summary of current findings around bilingualism and its impact on dementia. It also takes a look at some research that incorporated late bilinguals in their studies. The main contribution of this paper is that it offers suggestions for improvements of research methods moving forward, such as establishing a clear definition of the terms often used and generally increasing the effort put into this field of bilingual advantage.

Discussion

Introduction to the Bilingual Advantage Theory and Its Effect on Dementia:

Bilingual advantage influences several factors, including cognitive reserve (CR), which contributes to the delayed onset of dementia.[12] However, as previously stated, this theory is highly controversial. In order to understand the reason why this is the

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.15

case, it is important to first clearly state the definitions of bilingualism, its components, and the advantages.

What is the Bilingual Advantage? Bilingualism is the ability to speak and use two languages. The number of people with this skill is growing globally. Its benefits in communication skills and other linguistic terms are obvious, but there are non-linguistic advantages as well. In fact, bilingualism is believed to enhance executive functioning skills.[13] However, this theory, called the bilingual advantage,[7] is heavily debated due to the inconsistent and contradictory findings from different studies.

One of the executive functioning skills enhanced by bilingualism is inhibition, which is the ability to direct attention to one or a few important features of a stimulus while inhibiting attention to misleading or unnecessary features.[14] It is believed that bilinguals have stronger inhibition skills because of their constant need to suppress one language to focus on, better comprehend, and be fluent in another.[8] Bialystok *et al.* studied the inhibitory processes of bilinguals by assigning four trials of the Simon task to their 94 participants, where half of them were bilingual, and the other half were monolingual. They reported that young monolinguals had an average of 86.4% accuracy and 1,304 milliseconds reaction time, and that young bilinguals had an average of 97.1% accuracy and 536 milliseconds reaction time.[8] This trend was also true, and the gap was even greater, for older participants. This data shows that bilinguals performed better on the task than their monolingual counterparts, indicating that bilinguals have more efficient inhibitory processes than monolinguals.

Carlson and Meltzoff also studied inhibition in bilinguals compared to monolinguals.[14] They used an advanced version of Zelazo's Dimensional Change Card Sort (DCCS), which is a method that can assess the executive function of people of all ages.[15] Fifty kindergarten children were included in the study. Bilingual participants in the study scored 3.26 on the DCCS task, while the control group scored 2.12, a difference that is statistically significant. The difference between the scores indicates that bilinguals have better or improved executive function skills.[14] Since the Advanced DCCS requires inhibition to complete, it was evident that bilinguals have better inhibitory control than their monolingual peers. These two studies,[8,14] as well as others, support the hypothesis that bilingualism leads to better inhibitory control.

Another executive function that is often studied in this field is attention. This skill involves guiding attention to and focusing on an important aspect of a stimulus, which is in part made possible by inhibitory control, which suppresses irrelevant information.[16] Thus, if bilingualism boosts inhibitory control, then it would also benefit one's attention.

Task switching, the process of switching mental focus from one thing to another, is another factor of executive functioning that bilingualism can influence. Prior and MacWhinney studied whether cognitive bilingual advantage included these task-switching skills.[17] Bilingual and monolingual university students performed the task-switching paradigm, which was a task adapted from another study.[18] This task had three parts to it: single-task blocks (identifying the color or shape of a graphic), mixed-task blocks (identifying both color and shape), and single-task blocks again. In the mixed-task blocks, participants performed both switch (alternating between color and shape) and non-switch trials. As a result, the average reaction time for bilinguals was 437.97 milliseconds, and for monolinguals was 448.8 milliseconds for the single-task blocks. For mixed-block switch tasks, bilinguals' average reaction time was 814.16 milliseconds, while that of monolinguals was 875.54 milliseconds. The trend persisted for non-switch tasks. These data show that the bilingual participants in the study had faster reaction times than monolingual participants. Thus, it can be asserted that bilinguals have lower switching costs, i.e., there is a bilingual advantage in the efficiency of switching from one task to another. Researchers attribute this advantage to bilinguals' repeated practice of switching between languages.[17]

It is also said that bilingualism may benefit one's working memory, which is a temporary storage for information to be held in one's mind and used for cognitive tasks.[19] Based on the hypothesis that bilingualism affects executive functioning, Blom *et al.* studied the working memory of Turkish-Dutch bilingual and monolingual children from ages 4 to 6. They assigned the Dot Matrix, Odd-One-Out, and Forward and Backward Digit Recall tasks to children aged 6 and gathered their scores. Bilingual children' s mean scores were 16.1, 11.8, 20.0, and 7.5, respectively. Monolingual children's mean scores were 13.6, 9.9, 18.3, and 4.0, respectively. Comparing these scores shows the trend that bilingual children score higher in all tasks, indicating that they have better short-term memory. This indicates the presence of a bilingual advantage in working memory.[19] Another study also showed that loss of working memory due to old age was mitigated by bilingualism through an assessment of the Simon task.[8] These findings support the hypothesis that bilingualism impacts executive functioning.

Another component of executive functioning is cognitive reserve (CR). CR is a hypothesis that the brain has the ability to cope with damage. It maintains the brain's health and prevents cognitive decline by using efficient networks and alternative approaches to solve problems.[12] One of the main diseases that CR copes with is dementia, as it can delay the onset of the disease, and it is argued that life experience can impact the onset of dementia through contributing to CR.[12] Researchers are exploring how bilingualism may count as one of the experiences that delay dementia. For example, Craik and colleagues conducted research with the hypothesis that lifelong bilingualism contributes to CR. They collected data on the timing of onset and diagnosis of Alzheimer's disease (AD) from 211 monolingual and bilingual AD patients. Drawing from each patient's information about their occupational, educational, and language history, 102 patients were classified as bilingual and 109 as monolingual. Their reported data show that bilinguals reported the onset of AD symptoms 5.1 years later and were diagnosed with the disease 4.3 years later, compared to monolinguals. As a result, researchers concluded that lifelong bilingualism contributes to CR.[6] Combined with data from another study that found the same 5-year gap in bilingual and monolingual AD patients, this finding strengthens the CR hypothesis in terms of bilingual advantage.[5]

Although it seems convincing that bilingualism thus delays the onset of dementia, not all studies could replicate these findings. Chertkow *et al.* studied an even larger sample of 1842 monolingual, bilingual, and multilingual AD patients and found that the onset of AD was only 1 year later for bilinguals compared to monolinguals, a finding that is not statistically significant.[20] These conflicting findings make the bilingual advantage theory on dementia highly controversial.

Thus, some studies refute this claim,[9,20] but it can be asserted that there is at least some kind of bilingual advantage in terms of executive functioning, as well as delaying the onset of dementia. The extent of this advantage remains unclear. Also, even if it is supposed that there is an advantage, the findings of the studies that support this hypothesis are often inconsistent. There are a few factors that may help explain why these inconsistencies occur.

Factors that Define Bilingualism. One of the reasons why the bilingual advantage theory is debatable is that different studies use different definitions of bilingualism. Some researchers use surveys to assess and categorize their bilingual participants,[21] while others use tests for it.[22] The examples of factors of bilingualism include proficiency, differences, and the age of acquisition (AoA) of the languages.

It is believed that proficiency plays one of the largest roles in defining bilingualism.[23] Despite this, contrasting definitions of proficiency make the results of different studies incomparable. For instance, Zied *et al.* found that balanced bilinguals have better inhibitory control when compared to unbalanced bilinguals.[22] They measured this through the Stroop task,[24] where French- or Arabic-dominated bilinguals scored significantly lower than they did in the other language, while balanced bilinguals scored similarly on both languages. Balanced bilinguals are bilinguals who have similar levels of proficiency and daily use of both of their languages.[23]

A different study by Yow and Li on balanced bilingualism showed the same trend of advantage in executive functioning, using several tasks, including the Stroop task and the task-switching task.[21] But these findings are not necessarily replicates of each other since the researchers measured bilingual proficiency differently. Zied *et al.*[22] made their participants take the Boston Naming Test in order to measure their proficiency, while Yow and Li[21] conducted a survey where the participants self-reported their proficiency. These differences in the methods of measuring language proficiency make it hard to properly compare the findings, and thus, it leads to inconsistencies between studies.

Some studies[7,19,25] note the similarities and differences between the languages as an important factor for defining bilingualism. These similarities and differences are often defined by which linguistic family each language belongs to. For example, English and Spanish are in the same linguistic family in a broad sense, since both are in a big family called the Indo-European language family. In contrast, English and Turkish are in different families since Turkish belongs to a linguistic family called Turkic.

Petrosyan *et al.*[25] focused on these linguistic families in their study. They hypothesized that multilinguals who spoke similar languages outperformed those who spoke dissimilar languages in executive functioning. They found that among those with education, the former group of multilinguals scored 0.2, while the latter group scored 0.198. As a result, they concluded that multilinguals only outperformed monolinguals, and no major differences in executive functioning between different types of multilinguals were found.

Another study examined bilingual advantage in phonological awareness, which is strongly supported by executive functions.[7] Researchers had three groups of participants: Spanish-English bilinguals, English monolinguals, and Chinese-English bilinguals. Their interesting finding was that in the phoneme segmentation task, where participants were required to understand the phonetic segments of words, Spanish-English bilinguals performed the best out of all groups with a mean score of 0.59. English monolinguals came second with 0.48, and Chinese-English bilinguals performed most poorly with a mean score of 0.30. Since English and Spanish are of the same linguistic family, and English and Chinese are not, there may be some advantages for bilinguals who speak similar languages.

Adding on, a study by Blom and colleagues mentioned previously showed a bilingual advantage in executive functioning, especially working memory.[19] Since Turkish and Dutch are in different linguistic families, this finding might serve as evidence for better executive functioning in bilinguals who speak different kinds of languages, but no straightforward conclusion can be drawn because the study did not focus on the differences between different types of bilinguals. Thus, the answer to the question of whether similar or different languages lead to better executive functioning remains unclear.

Another factor that defines bilingualism is the age of acquisition (AoA) of the second language (L2). Different studies use different definitions of early bilingualism. For example, Soveri and her colleagues studied executive functioning skills in early bilinguals, who were, in this case, bilinguals who acquired their L2 before the age of 7. Task-switching paradigms and several other tasks showed that younger age, earlier L2 acquisition, combined with balanced use of languages, was associated with a smaller mixing cost. This can be an indicator that early L2 acquisition improves inhibition skills.[26]

In a different study, Tao *et al.* also studied executive functioning in bilinguals but with a different definition of early and late bilinguals. They stated that bilinguals who learned their L2 at or before age 6 were early bilinguals, and those who acquired L2 at or after age 12 were late. Similar to Soveri and colleagues,[26] they found that early bilinguals showed better executive functioning (in this case, monitoring processes) than late bilinguals.[27] Although their results showcase similar conclusions, their contrasting definitions of early bilinguals make it challenging to compare the two findings.

In 2019, De Bruin[10] published a review paper, communicating the need for improvements in research methods in the studies surrounding the bilingual advantage. She conveyed to other researchers in the field that, in order to facilitate further advancements in the field, there needs to be a standardized measurement of language proficiency. De Bruin also claimed that researchers need to provide a more detailed description of

DOI: 10.36838/IJHSR819.15

their participants. This will allow for the opportunity to eliminate the inconsistencies between the studies regarding the bilingual advantage. Furthermore, the factors of bilingualism can help determine which groups of bilinguals can enjoy this benefit. So, eliminating the inconsistencies will have a huge impact on the research field. Specifically, it can determine the benefits for late bilinguals.

How Late Acquisition of L2 Impacts the Onset of Dementia: It can be hypothesized that if early bilinguals have the bilingual advantage in delaying dementia, then late bilinguals may also have some advantage. The reason why this is an important topic to be discussed now is that among the studies that said dementia onset was delayed for bilinguals, most of them have focused on lifelong bilinguals, or bilinguals who have spoken both of their languages their whole life. As previously stated, Craik and his colleagues demonstrated in their research that bilinguals experienced Alzheimer's disease (AD) symptoms 5.1 years later than monolinguals, but they focused on the difference between lifelong bilinguals and monolinguals.6 Similarly, three other studies[5,16,28] determined that bilinguals reported AD symptoms around 4 years later than monolinguals. They also concluded that due to this, bilingualism may serve as a cognitive reserve (CR), which is linked to delaying dementia onset. But again, all of these studies only compared lifelong bilinguals to monolinguals.

Because of this, there is a limitation to the generalizability of these findings. Increased diversity in the types of participants is needed. Specifically, a focus on AoA is needed, as many bilinguals these days learn their L2 at different stages in their life.

Review of Studies with a Focus on Age of Acquisition. Despite what has just been stated, there are studies that compare early and late bilinguals. One example of this is research presented by Yow and Li[21] on executive functioning, specifically inhibition. As previously stated, through the Stroop task,[24] they reported that early and balanced bilinguals had better inhibition than late bilinguals. In a different study that studied executive functions' relation to CR, it was shown that inhibition is linked to CR,[29] which is then related to the delay of dementia onset.[12] Thus, Yow and Li's result means that early bilingualism benefits the onset of dementia more than late bilingualism does. But it is important to note that only bilinguals who acquired their L2 before the age of 7 were included in the study. So, the late bilinguals in this study may still be considered early bilinguals in other settings.

Another study that compared early and late bilinguals is by Grueso *et al.*[30] The participants of this research included three groups: early high-proficiency bilinguals, late high-proficiency bilinguals, and late low-proficiency bilinguals. Both the participants' proficiency and AoA were collected through a survey. Bilinguals were defined as early if they acquired their L2 before the age of 6, and late if after the age of 6. Through various tests, including the Cognitive Reserve Questionnaire, each participant's cognitive abilities were measured. Researchers found that high proficiency bilinguals outperformed those with low proficiency, and early AoA outperformed late AoA. Specifically, in the WAIS-IV coding test, which was used to measure participants' processing speed, participants with high proficiency performed similarly (around 0.00 to 0.15) regardless of their AoA, while participants with both late AoA and low proficiency performed relatively poorly (-0.30 to 0.00). This indicates that early high proficiency bilinguals had the best, and late low proficiency bilinguals had the poorest cognitive abilities out of the three groups. The researchers concluded that proficiency overrides AoA and that bilingualism, especially for those with high proficiency, may serve as a CR. Since CR may lead to a delayed onset of dementia,[12] this result indicates that earlier AoA and higher proficiency both contribute to delaying the onset of the disease.

Similar results were found in a study that focused on multilinguals instead of bilinguals,[31] and also in a study that compared early (AoA < age 10) and late (AoA > age 10) bilinguals in Canada.[32] In the former study, where researchers used multilinguals as their participants, participants were classified into either an active or a passive multilingual group with high or low proficiency, respectively. Information on their language history and executive functioning was collected, and the data show that active multilinguals scored 0.22 units higher in executive functioning than passive multilinguals, supporting the trend of increased bilingual advantage for multilinguals with high proficiency.[31] The latter study, which compared early and late bilinguals in Canada, also reported similar results. Using a questionnaire, participants were divided into early or late bilinguals with AoA described above. Their assessment of the flanker task showed that early bilinguals had the earliest response time of 557.9 milliseconds, while late bilinguals and monolinguals performed very similarly, with 569.8 milliseconds and 565.2 milliseconds response time, respectively. This indicates that early bilinguals with high proficiency have higher executive functioning.[32] However, the methods in which each study measured their participants are a bit different, so the results might only be comparable in some cases.

Another study comparing early and late bilinguals was done by Tao *et al.*,[27] but the results were somewhat contradictory to what has been discussed above. This study also focused on the question of AoA, balanced bilingualism, and the impact of these factors on executive functioning. Early bilinguals were defined as bilinguals with their AoA of the L2 at or before age 6, and late bilinguals were those with AoA at or after age 12. The level of proficiency was self-reported. The researchers gathered data through the Flanker task that early bilinguals outscored late bilinguals and monolinguals, and that late bilinguals performed better than monolinguals, indicating that any extent of bilingualism had a benefit on executive functioning over monolinguals. Adding on, and surprisingly, late bilinguals showed better ability in conflict resolution, which involves inhibition skills, than early bilinguals. The researchers thus concluded that different language experiences impact executive functioning differently. These results partially contradict what has been discussed above, indicating the fragility of this research topic and the importance and necessity for further research in the field.

The differences in methods, measures, and results between studies indicate that not all of the studies are comparable or consistent with one another. Due to this, the question of whether the AoA impacts the onset of dementia remains unclear. There is another topic closely related to this field, which is the comparison solely between different types of late bilinguals and their advantages in executive functioning.

Review of Studies with a Focus on Late Bilinguals. Several studies in the field only focus on late bilinguals and the differences within the population. One example of this is research that examined how late bi- and multilinguals' proficiency and daily use of different languages impact executive functioning, especially inhibition.[33] Their participants included late bilinguals and multilinguals who acquired one or more foreign languages at or after the age of 5. All factors, including daily use and proficiency, were measured through a questionnaire, and the results showed that higher mean proficiency was related to better inhibition. This indicates how proficiency might play a larger role than AoA or daily use in terms of its impact on executive functioning skills.

There was another study that focused on better executive functioning skills among late bilinguals, specifically brain plasticity. Lin[34] conducted research on late bilinguals in Beijing, China, to fill the gap regarding late bilinguals and how they may affect the delay of dementia. Since neuroplasticity is linked to CR[35] and thus the onset of dementia,[12] Lin questioned whether late AoA of L2 affects one's neuroplasticity. After comparing groups with different language exposure levels, she concluded that those with more exposure to English (192 hours) had better brain plasticity than those with less exposure (64 hours). These numbers of hours were equivalent to the number of academic years of learning English, where 192 hours meant 3 years and 64 hours meant 1 year. This result indicates how the time commitment to learning the L2 may also impact the onset of dementia. However, the researcher cites active lifestyle and social life responsibilities of the people in the researched area as factors that might have influenced these results, so the findings of this study may only be generalizable to people in Beijing, China. Since generalizability is crucial to determining the extent of bilingual advantage, this research is insufficient to answer the question of delaying the onset of dementia.

Another research paper that supports the bilingual advantage of late bilingualism focuses on the language control abilities of bilinguals. Festman and her colleagues[36] studied 29 late bilinguals with different language control abilities and measured the participants' executive functioning skills. Assessments through the Wechsler Adult Intelligence Scale[37] and several other executive function tasks showed that those with strong control abilities performed better on most of the tasks, especially in the WAIS tests, where they scored 17.2 on the information section, while those with weaker language control abilities only scored 13.4. The same trend persisted for all the other sections. The researchers concluded that participants with stronger language control ability have an advantage over those with weaker language control ability in terms of executive functioning. CR is a part of executive functioning, so better language control may play a role in the delay of dementia onset.

All of these different factors may indicate that late bilinguals also benefit from the bilingual advantage and the delay of dementia. But some studies refute the overall hypothesis. One of them is research done by Zahodne *et al.*[38] on late bilinguals who acquired their L2 in adulthood. Although their results supported that bilingualism improved basic executive functioning, they found that it was not connected to delaying cognitive decline and thus had no link to delaying dementia. They only included English-Spanish bilinguals as their participants, so their findings may not be generalizable to all, but the extent of generalizability is unclear. They concluded their research by hypothesizing that only bilinguals with higher education may exhibit an advantage in CR.

One explanation for this result differing from others may be the research methods used by researchers. In this study, researchers conducted a longitudinal study involving Cox regression, which helps examine the effect of multiple variables on the onset of a disease. In contrast, in Festman and colleagues' study, which indicated that late bilingualism may benefit executive functioning, the researchers mainly examined the data gathered from assigning multiple tests and tasks to their participants.[36] Since there are multiple papers mentioned previously that support these researchers' conclusion, the actual impact of bilingualism on the onset of dementia remains unclear.

Despite the need for further research, it can presumably be said that there is some kind of bilingual advantage that late bilinguals enjoy, whether it is the actual delay of dementia onset or some other factor of executive functioning. Due to the contradicting results, though, not all research in the field supports the hypothesis that any type of bilingualism has an advantage in dementia; further research on this question is needed in order to better define the extent of this proposed advantage.

Future Directions. No certain conclusion on the bilingual advantage for late AoAs can be drawn because of the inconsistencies derived from different definitions of bilingualism. These include measures of the AoA and the balancedness of bilingualism, as well as the method used to measure these factors (assessment or self-report). This showcases the need for a standard measurement of bilingualism, which is a part of the suggestions previously addressed by De Bruin in her broader review of bilingual advantage.[10] De Bruin proposed that the proficiency of the languages should have a standard measurement. Adding on to this, there should be a standard measure of AoA. This is the cutoff age where a bilingual person would be defined as early or late. One possible solution would be to determine the critical period of learning two languages simultaneously and to make the age at the end of that period a cutoff.

Furthermore, it is recommended to align the definition of L2 AoA so that it really reflects the AoA of that language. In other words, AoA should be the age when the person began using that language fluently and not the age when they started learning that language. Even if this was the intention for researchers using AoA as a factor in their research, they must

DOI: 10.36838/IJHSR819.15

make this clear to their participants, especially when collecting these data in a survey or some kind of self-report. This would allow the researchers to get their desired data on their participants. This definition of AoA was originally used by Luk and colleagues in their research on AoA's impact on cognitive control.[32] They used this definition because an individual cannot be fluent in a language that they have just started learning. This way of defining AoA should be used increasingly in future studies so that all studies have a consistent measure of AoA that aligns with the participants' onset of fluency in any language.

Also, there is a very limited amount of research that considers the bilingual advantage of early and late bilinguals who speak different kinds of languages. As mentioned previously, the similarities and differences between the languages people speak may have an impact on executive functioning. For this reason, it is important to take into account various factors of bilingualism in studies.

In general, more studies need to be conducted around late bilinguals and their advantage in dementia onset in order to arrive at a clearer conclusion. This is true with all the other types of bilinguals as well. These people would include, for example, bilinguals with low proficiency or those who learned their L2 in different settings, such as at school or at home. No specific suggestions can be made since this review does not focus on those groups of bilinguals. Going back to late bilinguals, if late bilingualism can be proven to be one of the ways that delays the onset of dementia to some extent, it will have a great impact on a global scale. By encouraging learning of an L2 for both children and older adults, we can aim to reduce the risk of dementia.[39,40] Even if the hypothesis becomes disproved, we can still encourage early L2 learning for reduced risk of dementia. Decreasing the number of dementia patients will not only improve the lives of each individual but also benefit the global economy. The cost of care for dementia is substantial for each patient, and that burden added up to 1.3 trillion USD globally in 2019, according to Wimo *et al.*[11] This cost is predicted to increase as time passes, so any effort to either prevent or delay dementia would be valuable.

Strengths and Limitations

There are several strengths to this review paper. To the author's knowledge, this is one of the first reviews that focuses on the relationship between late bilingualism and dementia. Another strength of this review is that it provides suggestions for further research that would contribute to clarifying the extent of the bilingual advantage.

As with all research, this review is not exempt from limitations. One limitation of this review is that it references existing literature that may not be generalizable. For example, Yijin Lin[34] conducted a study on the neuroplasticity of late adult bilinguals in Beijing, China. Although she found that a longer time commitment in language learning led to better neuroplasticity, the researcher mentioned that the findings might be influenced by social aspects of Beijing, so it is not generalizable. This limits the ability to review thoroughly and come to a clear conclusion.

Conclusion

Multiple studies indicate that bilingualism might impact the onset of dementia.[4–6] Among the studies that go in depth on bilingual advantage, many of them only take into account lifelong or early bilinguals.[6,14] Not so many of them put an emphasis on bilinguals who acquired their L2 later in their life.[21,32] This paper summarizes the current findings on the general bilingual advantage on the onset of dementia and then reviews papers that focused on late bilingualism and dementia.[34,36] In conclusion, the extent to which late bilingualism has an impact on the onset of dementia remains unclear and highly debatable.

In order to advance our knowledge about the bilingual advantage, a more detailed definition of bilingualism is needed, especially the AoA. This includes both establishing the standard age cutoff for early and late bilinguals, as well as determining which stage of language learning should be considered as the AoA. By improving the definition of AoA, the extent of late bilingualism's impact on the onset of dementia can be better examined. This will contribute to the possibility of decreasing the number of dementia patients because if this advantage for late bilinguals can be proven, then L2 learning can be encouraged at any stage of one's life. Even if future research disproves this advantage, L2 education from earlier on in life can be encouraged more. Reducing the number of dementia cases will decrease the hardship of both patients and caregivers, improve public health,[41] and benefit the global economy.[11]

Acknowledgments

The author would like to thank Dr. Jorge Avila for guidance and mentoring on this research project, as well as Bre Calhoun for helpful comments on the manuscript. The author attests that the ideas, graphics, and writing in this paper are entirely her own.

References

1. Wilson, D.; Peters, R.; Ritchie, K.; Ritchie, C. W. Latest Advances on Interventions That May Prevent, Delay or Ameliorate Dementia. *Ther. Adv. Chronic Dis.* **2011**, *2* (3), 161–173. https://doi.org/10.1177/2040622310397636.
2. Hall, C. B.; Lipton, R. B.; Sliwinski, M.; Katz, M. J.; Derby, C. A.; Verghese, J. Cognitive Activities Delay Onset of Memory Decline in Persons Who Develop Dementia. *Neurology* **2009**, *73* (5), 356–361. https://doi.org/10.1212/WNL.0b013e3181b04ae3.
3. Mangialasche, F.; Kivipelto, M.; Solomon, A.; Fratiglioni, L. Dementia Prevention: Current Epidemiological Evidence and Future Perspective. *Alzheimer's Res. Ther.* **2012**, *4* (1), 6. https://doi.org/10.1186/alzrt104.
4. Brini, S.; Sohrabi, H. R.; Hebert, J. J.; Forrest, M. R. L.; Laine, M.; Hämäläinen, H.; Karrasch, M.; Peiffer, J. J.; Martins, R. N.; Fairchild, T. J. Bilingualism Is Associated with a Delayed Onset of Dementia but Not with a Lower Risk of Developing It: A Systematic Review with Meta-Analyses. *Neuropsychol. Rev.* **2020**, *30* (1), 1–24. https://doi.org/10.1007/s11065-020-09426-8.
5. Perani, D.; Farsad, M.; Ballarini, T.; Lubian, F.; Malpetti, M.; Fracchetti, A.; Magnani, G.; March, A.; Abutalebi, J. The Impact of

Bilingualism on Brain Reserve and Metabolic Connectivity in Alzheimer's Dementia. *Proc. Natl. Acad. Sci.* **2017**, *114* (7), 1690–1695. https://doi.org/10.1073/pnas.1610909114.
6. Craik, F. I. M.; Bialystok, E.; Freedman, M. Delaying the Onset of Alzheimer's Disease. *Neurology* **2010**, *75* (19), 1726–1729. https://doi.org/10.1212/WNL.0b013e3181fc2a1c.
7. Bialystok, E.; Majumder, S.; Martin, M. M. Developing Phonological Awareness: Is There a Bilingual Advantage? *Appl. Psycholinguist.* **2003**, *24* (1), 27–44. https://doi.org/10.1017/S014271640300002X.
8. Bialystok, E.; Craik, F. I. M.; Klein, R.; Viswanathan, M. Bilingualism, Aging, and Cognitive Control: Evidence From the Simon Task. *Psychol. Aging* **2004**, *19* (2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290.
9. Paap, K. R.; Greenberg, Z. I. There Is No Coherent Evidence for a Bilingual Advantage in Executive Processing. *Cognit. Psychol.* **2013**, *66* (2), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002.
10. de Bruin, A. Not All Bilinguals Are the Same: A Call for More Detailed Assessments and Descriptions of Bilingual Experiences. *Behav. Sci*. **2019**, *9* (3), 33. https://doi.org/10.3390/bs9030033.
11. Wimo, A.; Seeher, K.; Cataldi, R.; Cyhlarova, E.; Dielemann, J. L.; Frisell, O.; Guerchet, M.; Jönsson, L.; Malaha, A. K.; Nichols, E.; Pedroza, P.; Prince, M.; Knapp, M.; Dua, T. The Worldwide Costs of Dementia in 2019. *Alzheimer's Dement.* **2023**, *19* (7), 2865–2873. https://doi.org/10.1002/alz.12901.
12. Stern, Y. Cognitive Reserve in Aging and Alzheimer's Disease. *Lancet Neurol*. **2012**, *11* (11), 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6.
13. van den Noort, M.; Struys, E.; Bosch, P. Individual Variation and the Bilingual Advantage—Factors That Modulate the Effect of Bilingualism on Cognitive Control and Cognitive Reserve. *Behav. Sci.* **2019**, *9* (12), 120. https://doi.org/10.3390/bs9120120.
14. Carlson, S. M.; Meltzoff, A. N. Bilingual Experience and Executive Functioning in Young Children. *Dev. Sci*. **2008**, *11* (2), 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x.
15. Zelazo, P. D. The Dimensional Change Card Sort (DCCS): A Method of Assessing Executive Function in Children. *Nat. Protoc.* **2006**, *1* (1), 297–301. https://doi.org/10.1038/nprot.2006.46.
16. Bialystok, E.; Craik, F. I. M. Freedman, M. Bilingualism as a Protection against the Onset of Symptoms of Dementia. *Neuropsychologia* **2007**, *45* (2), 459–464. https://doi.org/10.1016/j.neuropsychologia.2006.10.009.
17. Prior, A.; Macwhinney, B. A Bilingual Advantage in Task Switching. *Billing. Lang. Cogn*. **2010**, *13*, 253–262. https://doi.org/10.1017/S1366728909990526.
18. Rubin, O.; Meiran, N. On the Origins of the Task Mixing Cost in the Cuing Task-Switching Paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* **2005**, *31* (6), 1477–1491. https://doi.org/10.1037/0278-7393.31.6.1477.
19. Blom, E.; Küntay, A. C.; Messer, M.; Verhagen, J.; Leseman, P. The Benefits of Being Bilingual: Working Memory in Bilingual Turkish–Dutch Children. *J. Exp. Child Psychol*. **2014**, *128*, 105–119. https://doi.org/10.1016/j.jecp.2014.06.007.
20. Chertkow, H.; Whitehead, V.; Phillips, N.; Wolfson, C.; Atherton, J.; Bergman, H. Multilingualism (But Not Always Bilingualism) Delays the Onset of Alzheimer's Disease: Evidence From a Bilingual Community. *Alzheimer Dis. Assoc. Disord.* **2010**, *24* (2), 118–125. https://doi.org/10.1097/WAD.0b013e3181ca1221.
21. Yow, W. Q.; Li, X. Balanced Bilingualism and Early Age of Second Language Acquisition as the Underlying Mechanisms of a Bilingual Executive Control Advantage: Why Variations in Bilingual Experiences Matter. *Front. Psychol*. **2015**, *6*. https://doi.org/10.3389/fpsyg.2015.00164.
22. Zied, K. M.; Philippe, A.; Karine, P.; Valerie, H.-T.; Ghislaine, A.; Arnaud, R.; Didier, G. Bilingualism and Adult Differences in Inhibitory Mechanisms: Evidence from a Bilingual Stroop Task. *Brain Cogn*. **2004**, *54*, 254–256. https://doi.org/10.1016/j.bandc.2004.02.036.
23. Atkinson, A. L. Does Bilingualism Delay the Development of Dementia? *J. Eur. Psychol. Stud.* **2016**, *7* (1). https://doi.org/10.5334/jeps.375.
24. Stroop, J. R. Studies of Interference in Serial Verbal Reactions. *J. Exp. Psychol.* **1935**, *18* (6), 643–662. https://doi.org/10.1037/h0054651.
25. Petrosyan, S.; Strangmann, I. M.; Nichols, E.; Meijer, E.; Briceño, E. M.; Narayanan, S.; Lee, J.; Arce Rentería, M. The Association of Multilingualism with Diverse Language Families and Cognition among Adults with and without Education in India. *Neuropsychology* **2025**, *39* (3), 223–234. https://doi.org/10.1037/neu0000988.
26. Soveri, A.; Rodriguez-Fornells, A.; Laine, M. Is There a Relationship between Language Switching and Executive Functions in Bilingualism? Introducing a within-group analysis approach. *Front. Psychol.* **2011**, *2*. https://doi.org/10.3389/fpsyg.2011.00183.
27. Tao, L.; Marzecová, A.; Taft, M.; Asanowicz, D.; Wodniecka, Z. The Efficiency of Attentional Networks in Early and Late Bilinguals: The Role of Age of Acquisition. *Front. Psychol.* **2011**, 2. https://doi.org/10.3389/fpsyg.2011.00123.
28. Perani, D.; Abutalebi, J. Bilingualism, Dementia, Cognitive and Neural Reserve. *Curr. Opin. Neurol*. **2015**, *28* (6), 618–625. https://doi.org/10.1097/WCO.0000000000000267.
29. Roldán-Tapia, L.; García, J.; Cánovas, R.; León-Estrada, I. Cognitive Reserve, Age, and Their Relation to Attentional and Executive Functions. *Appl. Neuropsychol. Adult* **2012**, *19*, 2–8. https://doi.org/10.1080/09084282.2011.595458.
30. Grueso, S.; Sánchez-Benavides, G.; Santos-Santos, M.; Subirats, L.; Prados, F.; Grau-Rivera, O.; Calabria, M. Bilingualism and Cognition: The Impact of Age of Acquisition, Language Use and Proficiency in Healthy Older Adults. Research Square, November 14, 2025. https://doi.org/10.21203/rs.3.rs-7796112/v1.
31. Renteria, M. A.; Strangmann, I.; Liu, Y.; Khobragade, P.; Petrosyan, S.; Khandelwal, N.; Banerjee, J.; Nichols, E.; Meijer, E.; Narayanan, S.; Lynch, K. M.; Aksman, L. M.; Lee, J. Multilingual Profiles and Their Association with Cognition and Brain Structure among Older Indian Adults. *Alzheimer's Dement*. **2025**, *21* (S3), e099414. https://doi.org/10.1002/alz70857_099414.
32. Luk, G.; De Sa, E.; Bialystok, E. Is There a Relation between the Onset Age of Bilingualism and Enhancement of Cognitive Control? *Billing. Lang. Cogn*. **2011**, *14* (4), 588–595. https://doi.org/10.1017/S1366728911000010.
33. Boumeester, M.; Michel, M. C.; Fyndanis, V. Sequential Multilingualism and Cognitive Abilities: Preliminary Data on the Contribution of Language Proficiency and Use in Different Modalities. *Behav. Sci.* **2019**, *9* (9), 92. https://doi.org/10.3390/bs9090092.
34. Lin, Y. Learning A New Language at Older Ages: Neuro-Cognitive and Psychological Well-Being Effects. *Thesis PhD--Univ. Nottm.* **2021**.
35. Vance, D. E.; Roberson, A. J.; McGuinness, T. M.; Fazeli, P. L. How Neuroplasticity and Cognitive Reserve: Protect Cognitive Functioning. *J. Psychosoc. Nurs. Ment. Health Serv.* **2010**, *48* (4), 23–30. https://doi.org/10.3928/02793695-20100302-01.
36. Festman, J.; Rodriguez-Fornells, A.; Münte, T. F. Individual Differences in Control of Language Interference in Late Bilinguals Are Mainly Related to General Executive Abilities. *Behav. Brain Funct. BBF* **2010**, *6*, 5. https://doi.org/10.1186/1744-9081-6-5.

DOI: 10.36838/IJHSR819.15

37. Wechsler, D. *The Measurement of Adult Intelligence*; The measurement of adult intelligence; Williams & Wilkins Co: Baltimore, MD, US, 1939; pp ix, 226. https://doi.org/10.1037/10020-000.
38. Zahodne, L. B.; Schofield, P. W.; Farrell, M. T.; Stern, Y.; Manly, J. J. Bilingualism Does Not Alter Cognitive Decline or Dementia Risk among Spanish-Speaking Immigrants. *Neuropsychology* **2014**, *28* (2), 238–246. https://doi.org/10.1037/neu0000014.
39. Đolić, S. ENHANCING COGNITIVE RESERVE THROUGH FOREIGN LANGUAGE LEARNING: DELAYING ALZHEIMER'S DISEASE AND DEMENTIA IN THE AGING POPULATION. *Knowl. - Int. J.* 2025, *69* (5), 853–858.
40. Theobald, M. Can We Delay the Onset of Dementia by Studying English as a Second Language? **2020**, 51–64. https://cir.nii.ac.jp/crid/1050011771506810368.
41. Chapman, D. P.; Marshall Williams, S.; Strine, T. W.; Anda, R. F.; Moore, M. J. Dementia and Its Implications for Public Health. *Prev. Chronic. Dis.* **2006**, *3* (2), A34.

Author

Rio Imawaka is a high school student in California, United States. As a bilingual speaker of Japanese and English, she is interested in the impacts and benefits of her skills. She also has an interest in a broader field of psychology, which she hopes to explore more in college.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Treating Leukemia Using CRISPR-Cas Genome Editing to Improve Delivery of Allogeneic and Autologous Stem Cells

Rishabh A. Shah

Westlake High School, 4100 Westbank Dr, Austin, TX, 78746, USA; rishi.329.shah@gmail.com
Mentor: Dr. Hamidreza Shaye

ABSTRACT: Every 27 seconds, someone across the world is diagnosed with blood cancer. The global burden of leukemias and other blood cancers has increased despite interventional therapies such as chemotherapy, immunotherapy, and adult stem cell transplantation. This paper explores the use of precision biotechnologies such as the CRISPR-Cas gene-editing tool to improve patients' quality of life, refining the delivery method of regenerative blood stem cell therapies in leukemias. In allogeneic blood stem cell transplantation, CRISPR-Cas can facilitate immune tolerance to donor hematopoietic stem cells by cleaving parts of their Human Leukocyte Antigen (HLA) genes and by adding genes encoding protective proteins. This gene-editing technique thereby aids in transforming the limited blood stem cell donor pool to create a safer and more effective interventional therapy without requiring the close matching of HLA between the donor and the recipient. In contrast, it also functions to edit the patient's own blood stem cells in blood cell disorders like sickle cell anemias and thalassemias in order to provide long-lasting, disease-free survival. Current advances in genetic engineering technologies are poised to revolutionize the future of regenerative biomedicine, mobilizing the universal delivery of off-the-shelf stem cell products.

KEYWORDS: Biology, Oncogenetics, CRISPR-Cas, Leukemia, Stem Cell Transplantation (SCT).

Introduction

Global Burden of Leukemias:

Globally, leukemia is the second most common blood-related malignancy across all ages as of 2022.[1] Leukemia is the most common cancer among children less than 5 years old and has the highest mortality rate in early childhood.[2] In 2020, leukemias were the 11th most frequent cause of cancer-related mortality worldwide, causing more than 300,000 deaths.[2]

Defining Leukemia and the Subtypes of Leukemias:

Leukemia is a life-threatening blood cancer caused by the abnormal and uncontrolled proliferation of immature white blood cells. It is a cancer related to hematopoiesis, the production of blood cells, and exhibits diverse genetic characteristics.[3] Leukemias can be classified into subtypes based on the progenitor cell type, the speed of abnormal cell growth, and the stage of maturation arrest.[3] The major subtypes include: acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemia (CML).[4] Table 1 displays the characteristics of the four major subtypes of leukemias and provides additional information regarding them.[6-15] Of these subtypes, AML is the most common acute leukemia among adults and has the highest adult mortality rate among all leukemias. AML has an overall five-year survival rate of 32% for adult patients over 20 years and a five-year survival rate of 18% for elderly patients over 65 years, depicting the threat leukemia poses worldwide.[5] At the same time, ALL, the most common leukemia among children and adolescents, accounts for the highest mortality rate, 14%, in that age range.[2]

Table 1: Subtypes of leukemias with associated mutations, the most affected population, and the percentage proportion of all major leukemia subtypes; created in BioRender. A chunk of rarer leukemias, such as hairy cell leukemia and large granular lymphocytic leukemia, are omitted; hence, the ranges in the "Percentage of leukemias" column do not add up to 100%. Breakpoint Cluster Region - Abelson murine leukemia viral oncogene homolog 1 fusion gene (BCR-ABL1). Philadelphia chromosome (Ph+). Fms-like tyrosine kinase 3 (FLT3). Nucleophosmin 1 (NPM1). Tumor Protein P53 (TP53). Short arm ("p") of chromosome 17 (17p). Data were acquired from a variety of sources.[6-15]

Subtypes of Leukemias

Type	Mutations associated	Primary population	Percentage of all leukemias	References
ALL	BCR-ABL1 (Ph +)	Pediatric	10-12%	[6 - 15]
AML	FLT3, NPM1, BCR-ABL1 (Ph +)	Adults	31-35%	[6 - 15]
CLL	TP53, 17p	Adults	33-38%	[6 - 15]
CML	BCR-ABL1 gene (Ph+)	Adults	15%	[6 - 15]

Etiology and Pathophysiology of Leukemia:

Often, leukemias can be caused by a variety of factors, including genetic mutations and environmental factors.[4] Genetic mutations can be either inherited via syndromes such as Down syndrome or Li Fraumeni syndrome, or acquired from unsafe exposure to radiation or chemicals. An example of this is the BCR-ABL fusion gene, which arises from an acquired mutation from a chromosomal translocation identified as the Philadelphia chromosome (Ph), ultimately leading to the development of CML and AML. Environmental factors that

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.16

may cause leukemia include ionizing radiation from radiation therapy or nuclear accidents, exposure to cigarette smoke, exposure to chemicals such as benzene and formaldehyde, and viruses such as Human T-Lymphotropic Virus-1 (HTLV-1) and Epstein-Barr virus (EBV).[16]

Leukemic stem cells (LSCs) are abnormally mutated hematopoietic stem cells in the bone marrow. LSCs are a critical component of leukemia because they trigger the onset of AML, causing uncontrolled cancer cell growth and thereby increasing their intrinsic resistance to chemotherapy.[17] The malignant transformation of hematopoietic progenitor cells (HPCs) can lead to an abnormal accumulation of immature cells in the bone marrow. This process causes an overflow of white blood cells and interferes with regular blood cell production.[16] In addition, Interleukin-1 (IL-1), a group of cytokines (signaling proteins), is overproduced in AML, while normally they are underproduced. Interleukin-1 receptor accessory protein (IL-1RAP) overproduction has been traced to myeloid malignancies in its signaling pathway, and it is involved in processes such as rapid cancer cell proliferation and leukemic cell shielding, which can further promote the progression of AML.[18] Hence, to limit the progression, it is imperative to target IL-1RAP to rupture those downstream pathways that would otherwise lead to cancerous transformation.[18]

With these multiple mechanisms of complex interactions triggering mutations and the proliferation of immature white blood cells, the body's primary immune defense becomes compromised and dysfunctional. Along with it, overcrowding in the bone marrow, as seen in Figure 1, can lead to the suppressed production of other cell lines like red blood cells and platelets, leading to anemia and clotting interference, respectively.[3,19] These pathological processes may result in symptoms such as fever, fatigue, pallor, bone pain, bleeding and easy bruising, recurrent infections, and lymph gland enlargement, among many others, as seen in Figure 2.[20]

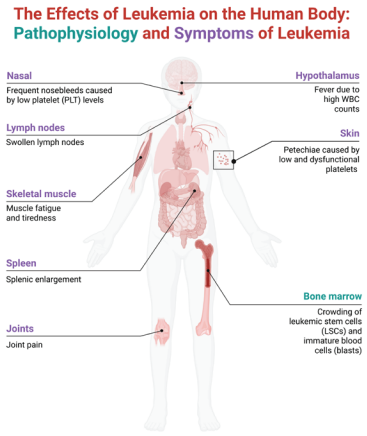


Figure 1: A portrayal of the "crowding out" aspect of leukemia in which a rapid increase in the amount of blasts and leukemic stem cells (LSCs) in the bone marrow can lead to implications for all three blood cell lines as listed above. This figure was made in BioRender and adapted from an image by Hamid, 2013.[19]

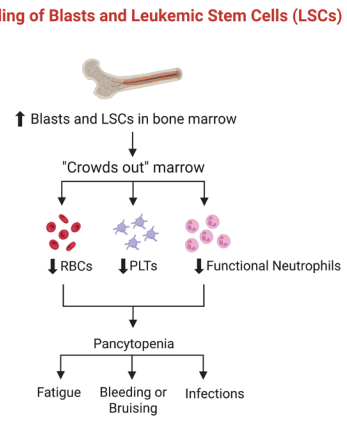


Figure 2: A visual depiction of the pathophysiology and common symptoms of leukemia. This figure was made in BioRender, and the information was gathered from Khwaja *et al.*, 2016.[20]

Introduction to Leukemia Therapies:

Current leukemia treatments include standard interventional therapies such as chemotherapy, immunotherapy, and radiation therapy, as well as advanced therapies like CAR-T cell therapy and stem cell transplantation.[21] These therapies are discussed in more depth later in this paper, focusing both on their clinical effectiveness and the adverse events that can impact patients' quality of life. Stem cell transplantation, the Human Leukocyte Antigen (HLA), and CRISPR-Cas9 genetic engineering in stem cell transplantation are introduced below.

Stem Cell Transplantation: Advanced Therapeutic Potential in Leukemias:

Stem cells are a subtype of self-renewing cells in our body tissues, with the ability to differentiate into specific cell lines: hematopoietic (blood-producing), mesenchymal (resembling embryonic connective tissue) cells, and others.[17] For this review, we will restrict our discussion to hematopoietic stem cells (HSCs), which can regenerate new healthy blood cells and replace the existing diseased cells. This ability of HSCs makes them an effective therapeutic option in leukemias and blood cell disorders such as hemoglobinopathies.

Allogeneic (allo-SCT) and Autologous (auto-SCT) Stem Cell Transplantation:

Stem cell transplantation is a cellular therapy that can potentially restore damaged tissue through its regenerative properties. There are two subtypes of stem cell transplantation—autologous (auto-SCT), consisting of the individual's own stem cells, and allogeneic (allo-SCT), where stem cells from a matched donor are obtained. These matched donors may be a family member or unrelated to the patient. While auto-SCT is more accessible and timely, it does pose a risk of relapsed leukemia due to possible residual disease, thereby limiting progression-free survival.[22] Allo-SCT is notably more effective in treating leukemia in contrast to auto-SCT.[22,23] However, it requires close matching of the Human Leukocyte Antigen (HLA), an immune-recognition marker of self versus foreign, between the patient and the donor. HLA compatibility helps prevent the rejection of transplanted tissue as well as re-

duces the probability of an adverse immune response mounted by the transplanted tissue against the patient's normal tissues, known as graft-versus-host disease (GVHD).[24] Despite these benefits, it is critical to recognize that the process of securing an HLA-matched donor can be time-consuming and burdensome.

Human Leukocyte Antigen, Immune Tolerance, and Allo-SCT:

HLA proteins are surface markers on leukocytes (white blood cells) and are encoded by Major Histocompatibility Complex (MHC) molecules, which are critical components of our immune system. They are Class I and Class II genes located on the short arm of chromosome 6.[25] There are three loci for Class I (HLA-A, HLA-B, and HLA-C); each individual inherits A, B, and C from each parent, resulting in six allelic pairs in each cell.[25] Whereas Class II molecules are mainly expressed on immune cells, especially those able to present foreign proteins (antigens). They have three loci (HLA-DR, HLA-DQ, and HLA-DP), of which HLA-DR is the most prominent and polymorphic subtype. HLA proteins are responsible for initiating immune rejection of foreign pathogens from their position on the surface of cells, because of their ability to differentiate self from non-self.[26] It is imperative to note that HLAs align more closely with other HLAs of the same family and next with those of the same ethnicity, due to the unique set of genes that comprise them. Each ethnicity presents its own unique HLA profile created by varying frequencies of inherited alleles.[27] Allo-SCT requires close HLA matching (the process of identifying and matching all HLA pairs) to prevent rejection of the transplant and GVHD. The need for HLA typing and matching is currently a huge barrier to access to stem cell therapy in leukemia.

Gene Editing Biotechnologies: TALENs, ZFNs, and CRISPR-Cas:

The focus of this review is on the utilization of scientific advancements in precision medicine biotechnologies like transcription activator-like effector nucleases (TALENs), zinc-finger nucleases (ZFNs), and CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats - Cas endonuclease) for the timely and efficient delivery of stem cell therapies.[28] The revolutionary potential of such gene-editing techniques, which are evolving to fix gene mutations and add functional genes *ex vivo*, has redefined both transplantation and regenerative.[29]

CRISPR-Cas: the Present and the Future:

Genetic editing biotechnologies, especially the more precise CRISPR-Cas system, have shown promise to make allo-SCT more universal, precluding the need to closely match the HLA of the donor with that of the patients. This gene modification tool uses the HLA-abrogation process, consisting of the disruption and self-repair of stem cell DNA.[30] This precision delivery technique would allow allo-SCT regimens with superior effectiveness to induce prolonged remission of leukemias while also being available relatively quickly. Additionally, this technology will help in minimizing the likelihood of adverse events such as GVHD.[31] Moreover, incorporating CRISPR-Cas genome editing to repeal HLA-mediated immune rejection will limit the need for immunosuppressive drugs with serious side effects.[32] CRISPR-Cas gene editing has also shown great success as a therapeutic approach in the improved delivery of autologous stem cells with the utilization of *ex vivo* gene editing to "silence" or "reactivate" certain responsible target genes.[33] This aspect of CRISPR-Cas gene editing has been elaborated in the discussion section of this paper.

Gene Editing, Stem Cell Transplantation, and Hemoglobinopathies:

Genetically inherited red blood cell disorders such as sickle cell disease and beta-thalassemia major are the most common inherited diseases around the world, affecting 7% of the world's population.[34] Although non-cancerous, their global fatality at a staggering 376,000 lives lost annually positions them as a major global health issue.[34] The overall disease burden of these conditions makes it critical to work on prevention and therapies. CRISPR-Cas9 gene-editing technology has promise in enabling long-lasting disease-free survival for patients with these blood cell disorders through edited autologous stem cell transplantation.

Discussion

Current Therapies in Leukemia:

Current interventional therapies for blood cancers include chemotherapy, radiation therapy, targeted therapies such as tyrosine-kinase inhibitors, novel immunotherapies such as immune checkpoint inhibitors, natural killer-mediated immunotherapy, and chimeric antigen receptor (CAR) T-cell therapy, and monoclonal antibodies. These therapies have indications, advantages, and limitations, such as adverse effects that may cause poor quality of life or failure of treatment itself. Hence, the usage of stem cell transplantation as a therapeutic approach for blood cancers becomes essential in refractory and relapsed leukemias. The characteristics of and statistics related to these therapies are displayed in Table 2.[47-50]

Chemotherapy:

Chemotherapy is the most widely used standard of care treatment protocol for leukemias.[3] Chemotherapy is a therapeutic cancer intervention that causes genotoxicity to damage cellular DNA.[21] The chemical agents produce reactive oxygen species—unstable and reactive molecules, which have the potential to cause oxidative stress and induce cellular damage to the cancerous cells for the desired effect. Unfortunately, the administration of chemotherapy also damages healthy tissues, inducing a series of unwanted short-term adverse effects such as severe fatigue, nausea, vomiting, diarrhea, headaches, and alopecia, as well as long-term side effects such as organ damage, secondary malignancies (e.g., myelodysplastic syndromes), and neurotoxicity.[35]

DOI: 10.36838/IJHSR819.16

Targeted Therapy and Immunotherapy:

The second arm of intervention in leukemias consists of immunotherapies, which can be systemic and targeted as well. Immunotherapies, often used alongside chemotherapy, are a form of therapeutic cancer intervention in which the body's immune cells and antibodies are used to eliminate pathogenic cells.[36] The most common immunotherapies are chimeric antigen receptor (CAR) T-cell therapy, natural killer-mediated immunotherapy with specific targeted mechanisms, and monoclonal antibodies, including immune checkpoint inhibitors, whereas the targeted therapy with tyrosine kinase inhibitors is not considered to be an immunotherapy.[36]

Chimeric antigen receptor (CAR) T-cell therapy is a targeted immunotherapy that works by attaching a ligand to a receptor for the purposes of disrupting cancer-causing processes, and then it is able to target specific cells with toxic components.[21] The efficacy of CAR T-cell therapy is notably threatened by a phenomenon known as "low-antigen escape," which means the target is not immunogenic enough to be destroyed by the T-cells delivered.[37] It can be affected by the on-target, off-tumor toxicity as well, highlighting major deficiencies in utilizing CAR T-cell therapy in leukemic interventions.

The other forms of immunotherapies are natural killer (NK)-mediated immunotherapies, which are meaningful as they have the ability to detect cells that lack MHC class I molecules. This trait could allow the potential pairing of the usage of natural killer (NK)-mediated immunotherapies and CRISPR-Cas genome editing for targeted killing of cancerous cells that remain after allo-SCT. But their use can be limited due to their potential to induce hyperstimulated activation and thereby weaken the cytotoxic functions of natural killer (NK) cells.[38,39] This functional deficiency hinders NK cells from performing their normal immunoregulatory functions, thereby predisposing the body to infections. This leads to a potential path to study a combination of therapeutic interventions in blood cancers, in order to make up for the limitations of one therapy with the advantages of another.

Monoclonal antibodies (mAbs) are another form of targeted systemic immunotherapy in leukemias. The mAbs are glycoproteins targeting specific surface antigens and molecules on tumor cells.[36] They are divided into four major categories: murine (omab), chimeric (xiomab), humanized (zumab), and fully human (umab) mAbs.[36] Nearly half of all antibodies are used in therapeutic cancer interventions.[40] MAbs have a variety of methods by which they function, including blocking receptor-ligand binding and inhibiting cell proliferation. However, the lack of cost-effectiveness makes therapeutic antibodies unconventional, as they limit access to a wider patient base.[36]

Immune checkpoint inhibitors (ICIs) are a distinct type of monoclonal antibodies (mAbs) that regulate immune cell activation for an anti-tumor response and combat the self-immunity of certain cancer cell types. These cancer cells would otherwise evade destruction by exploiting the body's natural immune checkpoints of preventing autoimmune attack.[41] ICIs provide a more promising and specialized therapeutic intervention, yet their vulnerability to immune escape by leukemia cells and the adverse events from possible autoimmune attack against the body's own tissues pose an obstacle to their incorporation.[41]

Tyrosine kinase inhibitors (TKIs) are small molecules that target and disrupt signaling pathways that allow malignant cell proliferation.[42] Molecular changes made by TKIs can lead to cell death via various pathways.[42] They have been used for the last twenty-five years and have been the cornerstone of management of chronic myeloid and chronic lymphocytic leukemias, which are the hematologic malignancies heavily dependent on tyrosine kinases for cellular proliferation and survival. But many of these TKIs are not selective and may inhibit multiple kinases, thereby causing adverse events related to organ damage from "off-target" effects. The limiting factor can also be the adverse events resulting from the "on-target" effects due to excessive inhibition of the tyrosine kinases. These can be in the form of gastrointestinal, cardiac, and skin complications, which can cause short-term or long-term damage.

Radiation Therapy:

Radiation therapy (RT) is another therapeutic intervention specifically utilized for certain types of leukemias like ALL, albeit it is a less frequently indicated treatment for leukemias overall. RT utilizes high-energy beams to damage cancerous cells in leukemia. RT has been used as a bridge therapy to complement alternative cancer interventions such as chimeric antigen receptor (CAR) T cell therapy in certain leukemias, such as ALL, where it sensitizes tumor cells to be attacked by T-cells.[43] While it is effective in damaging cancerous cells, it can unfortunately cause an adverse reaction to healthy tissues as well, creating a paradoxical response. For example, in CAR-T cell therapy, if the RT dose is overtly robust, it can induce early activation of T-cells, thereby causing destruction of those very immune cells that were supposed to protect against cancerous cells.[43] This has become a limitation in incorporating RT during therapeutic interventions.

Stem Cell Transplantation: Advanced Therapeutic Potential in Leukemias:

Table 2: Efficacy is measured by the 5-year OS of various therapies for ALL. The table below, created in BioRender, reviews the leading therapies for leukemias, including chemotherapy, immunotherapy, radiation therapy, stem cell transplantation, and targeted therapy. While ALL is most common in pediatric populations, data for immunotherapy usage in ALL were limited to adult populations due to the unavailability of pediatric populations. Overall survival (OS). Leukemia-free survival (LFS). Data were acquired from a variety of sources.[47-50]

Current Therapies to Treat Acute Lymphoblastic Leukemia (ALL)

Type of Therapy	Mechanism of Action	Advantages	Disadvantages	Data collection population	Efficacy (5-year OFS)	References
Chemotherapy	DNA damage induced by cytotoxic drugs	Rapid drug onset	Healthy cell damage, drug resistance, dosage selection difficulty	Pediatric	85-90%	[47 - 50]
Immunotherapy	Manipulation of immune system to facilitate cancer cell destruction	Sustained anti-tumor effects	Immune related adverse events (irAEs), hyper-progression, reactivation of disease	Adults	65-70% (using 67.8% 5-year LFS)	[47 - 50]
Radiation Therapy	DNA damage caused by X-rays	Localized control of cancer	Tissue damage, low blood counts	Pediatric	70-80%	[47 - 50]
Stem Cell Transplantation (SCT)	Replacement of dysfunctional bone marrow with healthy HSCs	Tissue regeneration, avoidance of invasive surgery	Tumorigenesis, graft versus host disease (GVHD), drug toxicity	Pediatric	70-90%	[47 - 50]
Targeted Therapy (TKIs, venetoclax, etc.)	Targeting of specific molecules to slow cell growth	High specificity, reduced toxicity, oral administration	Treatment resistance, dermatologic adverse events (dAEs)	Pediatric	>80%	[47 - 50]

Given the limitations of the aforementioned therapies for leukemias, it is key to explore the treatment modality of stem cell transplantation (SCT) as one of the leading alternatives, especially in refractory and relapsed cases, as well as for those cases with an adverse cytogenetic profile.[44,45] Stem cell transplantation offers the potential for long-lasting remission as well as cure in leukemias, making it the second-line treatment for certain refractory and relapsed cases of acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL).[46] "The use of bone marrow transplantation (a type of SCT) from compatible donors is the most effective treatment for CML" because alternative therapies might not work effectively and cause severe side effects to patients.[46]

Autologous and Allogeneic SCT as Therapeutic Interventions in Leukemias, Hemoglobinopathies, and Inherited Red Blood Cell Disorders:

There are two distinct types of stem cell treatment based on their source: autologous and allogeneic. Autologous stem cell transplantation (auto-SCT) consists of transplanting the individual's own stem cells, in contrast to allogeneic stem cell transplantation (allo-SCT), where the stem cells from a matched donor are obtained from either a family member or an individual unrelated to the patient. Their indications of use in leukemia are based on the patient profile, type of leukemia, and the response to available therapies.[46]

In autologous stem cell transplant therapy, the patient's intrinsic stem cells are collected and treated *ex vivo* before reinfusion into the patient's own body to treat a condition. This can only be done if their bone marrow is healthy enough to obtain stem cells. Auto-SCT is more accessible, timely, and has the advantage of better acceptance of transplanted tissue without the need for any HLA matching. But auto-SCT does pose a risk of relapsed leukemia due to possible residual disease, thereby limiting the progression-free survival of affected patients.[22] Hence, for its effective use therapeutically to treat the disease condition, editing with genetic engineering technologies like the CRISPR-Cas9 endonuclease system can facilitate "silencing" or "reactivating" a gene to make it functional. Thereafter, these edited intrinsic stem cells can be reinfused in the patient to treat their leukemia, which is considered an autologous transplant.[33] While this process of "cleaning out the residual disease from harvested stem cells using gene edits" is still undergoing clinical trials for auto-SCT in leukemias, it has been successfully carried out for patients with inherited red blood cell disorders—hemoglobinopathies like sickle cell disease. An example is Casgevy, which is the first FDA-approved CRISPR-based treatment for severe sickle cell disease and beta-thalassemia; this treatment has been successful in clinical use as of 2023. The disadvantage of incorporating autologous stem cell transplantation in the therapeutic protocol for leukemia is that the intrinsic stem cells could be cancerous in the affected individuals with leukemia, thereby potentially causing relapse of the disease, while not helping replenish hematopoietic (blood-producing) cells easily.[51] The other major limitation is the cost related to this modality of treatment, thereby creating inequitable access.

An allo-SCT is the transplantation of stem cells from another individual—related or unrelated—being transfused into the patient's body to treat the blood disorder. Such transplantation needs to be matched for HLA, the immune-recognition marker, and offers the advantage of being more efficacious in inducing remission due to the graft-versus-leukemia (GVL) effect. This makes it a more sought-after treatment of certain leukemias in comparison to autologous.[24,46] Allo-SCTs may facilitate longer remission and prevent progression of the disease, depending on the specific leukemia and other individual risk factors.[22,23] However, there are limiting factors for allogeneic stem cell transplantation since close compatibility of the donor's HLA with the patient's is necessary for a successful transplant outcome. It is also imperative to align HLA closely to help prevent rejection of the transplanted tissue for ameliorating the affected patient's condition. This would also help minimize the risk of an adverse immune response mounted by the transplanted donor tissue against the patient's normal tissues, graft-versus-host disease (GVHD).[24,51] But as evidence shows, the process of securing a fully matched stem cell donor can be labor-intensive and time-consuming, which can delay the chances of a patient's medical condition improving. In the meantime, affected patients can continue to worsen and possibly suffer complications from their disease. It takes an average of 52 days to get to the matched donor information session for beginning the process of allo-SCT, which is longer than the average duration of an auto-SCT.[52]

This paper investigates the progress made in the precision delivery of effective allogeneic stem cells using the CRISPR-Cas system genome editing to modify HLA genes and to facilitate successful acceptance by the recipient.[53] It is worthwhile to consider the potential success of genetic engineering techniques like zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and CRISPR-Cas genome editing in precision medicine for abrogating HLA genes. These HLA-deficient or HLA-negative donor stem cells or lymphocytes become hypoimmunogenic, thereby evading the host immune system's detection.[26,54] The resulting successful allogeneic stem cell transplants potentially induce a durable remission of the leukemia and lead to event-free survival in the affected patients.

Gene Editing Biotechnologies: TALENs, ZFNs, and CRISPR-Cas:

First, we delve into the uses and limitations of ZFNs and TALENs in this realm. Then, we will discuss the advantages and detailed processes of utilizing the CRISPR-Cas genome editing tool to help in rendering "universal" donor stem cells. ZFNs and TALENs are programmable nucleases that consist of a nonspecific DNA cleavage portion linked to sequence-specific DNA-binding modules.[29] They induce these edits in the genes by making double-stranded breaks in the DNA, which thereafter stimulate DNA repair mechanisms in the form of homology-directed repair (HDR) at specific genomic locations or non-homologous end-joining (NHEJ); these mechanisms are also called "gene-knockins" and "gene knockouts", respectively.[54] Although the flexibility of this targeted

DOI: 10.36838/IJHSR819.16

approach has brought genome engineering to the forefront of therapeutics in leukemias, they are more time-consuming, costly to manufacture, and lack site-specificity at times. This is where more efficient delivery systems like the CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats - Cas endonuclease) gene editing system have been a better alternative comparatively for targeted genetic modifications.[29]

Introduction of the CRISPR-Cas Gene Editing Tool:

CRISPR-Cas9 is an innovative gene editing tool functioning as a pair of molecular scissors that utilizes a single-guide RNA (sgRNA) and Cas9 protein duplex to edit DNA.[30] This is an important part of the adaptive immune system in prokaryotes to direct the precise removal or insertion of DNA fragments with a bacterial enzyme. The bacterial CRISPR locus was first described by Francisco Mojica, and it is made of viral or plasmid DNA that previously infected the microbes, found between the array of short palindromic sequences, called "spacers."[55] It was patented just over a decade ago by E. Charpentier and J.A. Doudna, who repurposed the type II CRISPR/Cas9 system for use in therapeutics. This gene editing tool has its roots in an adaptive bacterial immune system—bacteria loading a portion of the attacking viral DNA into their own genome as "memory" for the future and for prevention of re-infection with the virus. Target recognition by the Cas9 protein, an endonuclease ("the molecular scissors"), requires an identifying sequence known as protospacer adjacent motif (PAM), "the postal address," which is upstream of the cRNA-binding region. This allows the CRISPR-Cas complex to be adaptive as per the necessities of gene editing by redesigning the crRNA to cleave any DNA sequence.[56] Compared to ZFNs and TALENs, it is more effective and simpler to produce with a short turnaround time, thereby reducing the economic burdens involved.[24,29] The different delivery methods for CRISPR-Cas9 include viral gene delivery (e.g., adenovirus, lentivirus, retrovirus) and non-viral gene delivery systems such as DNA electroporation, transposons, nanocarriers, and cell-squeeze techniques involving various bioactive materials.[29] The details of these delivery systems, including uses and limitations, are beyond the scope of this paper.

CRISPR-Cas in Stem Cell Transplantation: Advantages and Limitations:

CRISPR has evolved to help delete genes of HLA Class I and II antigens, which make the donor stem cells "escape" the patient's immune system, becoming more acceptable to the patient's body; Figure 3 depicts this HLA cleavage and its intended effect.[57] As the rejection of foreign stem cells can be prevented, it precludes the need for closely matching the HLA as well.[30,32,54] This facilitates such stem cells from the donor to become more "universal" in nature and ready "off-the-shelf" for the purpose of effectively treating rare and refractory chemo-radiation-resistant leukemias while minimizing the wait time to find HLA-matched donors. The additional uses of CRISPR in the arena of stem cell transplantation are the prevention of automatic destruction of such HLA-deficient stem cells by the natural killer (NK)-cell response, which does this by overexpressing NK-inhibitory ligands.[38,39] The other method where CRISPR is useful in allogeneic stem cell transplants is the T-cell receptor "gene-deletion" in donor T-lymphocytes, which can prevent them from launching an attack on the recipient's foreign cells, thereby preventing graft-versus-host disease and related adverse events.

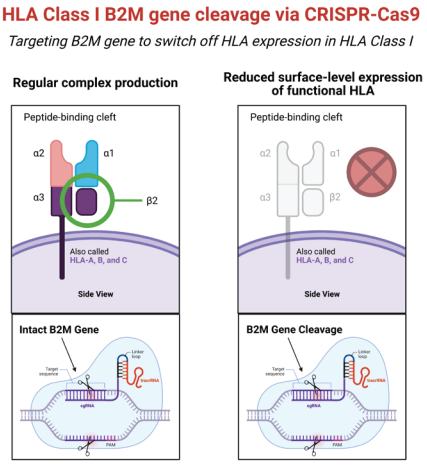


Figure 3: A visual depiction of the potential to reduce surface-level expression of functional Class I HLA by targeting the beta-2-microglobulin gene via CRISPR-Cas9. This figure was made in BioRender and adapted from a figure made in collaboration between BioRender and the Doudna lab.[57]

The CRISPR-Cas system has provided a quicker and simpler solution to manipulate the genome for therapeutic benefits, but it still has several limitations that need to be evaluated further and alleviated.[30] Just like we need to discuss the indications and advantages of the CRISPR-Cas system, we need to note the limitations of this gene-editing tool as well. The limitations can be the possibility of tumorigenesis through triggering of a p53-dependent DNA damage response that induces cell death and apoptosis of such affected cells with functional p53, whereas the ones with p53 mutations get naturally filtered out for survival and uncontrolled multiplication.[29] This can result in the suppression of the gene-editing tool itself due to the arrest of the cell cycle and division, which prevents CRISPR-related repair mechanisms from being successful.[29] There are also chances of genome editing by CRISPR being rendered ineffective due to an adaptive immune reaction against it by the host's immune system. In addition, there is a higher risk of off-target mutations due to genomic instability and unintentional disruption of other genes, which may need a "switch-off" mechanism. There is promise in single-base editing or substitution to minimize the off-target effects, as well as disruption of genes, causing inadvertent mutations during repair.[58]

■ Conclusion

Hematopoietic stem cell transplantation (HSCT) has largely been successful in treating certain leukemias, the ones that are refractory or resistant to conventional therapeutics, and in cases where those have caused severe adverse effects for the patients. Data shows although the disease burden of leukemias has increased recently, the good news is that overall survival (OS) and event-free survival (EFS) across all leukemias and

patient populations, especially in the pediatric population, have also increased.[2] We can certainly credit these to novel and safer targeted therapies, immunotherapeutics, and stem cell transplantation in combination with precision medicine and gene-editing technologies. Of the current genetic engineering tools, ZFNs and TALENs have some limitations due to their off-target effects in their practical applications, which makes the CRISPR-Cas system more viable as a therapeutic tool for effective and more specifically targeted delivery in leukemias. Stem cell transplantation, especially allogeneic, has more potential for enduring remission in patients with refractory, relapsed leukemias as compared to autologous, which has limited indications in certain acute myeloid leukemias. The need for close HLA compatibility in allogeneic stem cell transplants for safe and successful transplant outcomes makes it extremely unlikely for many patients to access the limited stem cell donor pool over their lifetime. Their quality of life and survival chances are adversely affected. Hence, the use of gene-editing techniques, especially CRISPR-Cas duplexes to deliver hypoimmunogenic and "universal" allogeneic stem cells in a quicker turnaround time to patients with leukemias, seems extremely promising.

The goal of therapeutic gene editing is to make potentially curative stem cells safer by reducing the off-target effects and associated risks to patients.[56] These processes are still evolving, with better design of the CRISPR-Cas guide RNA, using a variant of Cas9 nickase (Cas9n) for inducing single-stranded breaks, as well as with single-base and prime editing; this should help better accuracy, obviating the off-target effects. Additionally, the newer technologies of direct *in vivo* gene-editing tools explored in other genetic musculoskeletal and solid tissue tumors can be extrapolated to therapeutic applications in blood cancers and blood cell disorders. However there are some limitations of genetic engineering technologies, like the economic burdens on the healthcare system, differential access to resources in varied settings, and the need for solutions in order to reduce such inequities created from the very high costs of CRISPR-edited stem cells. Moreover, it is necessary to adhere to the ethical regulatory frameworks of the government and entities about genetic modification that may inadvertently cause long-lasting complications in recipients, including tumorigenesis.[59]

In summary, gene-editing precision tools like CRISPR-Cas have kick-started a revolutionary transformation in the treatment options for leukemias and, in the future, have the potential to lead to improved outcomes of better quality of life for such patients in a more equitable manner.

AI Acknowledgments

During the preparation of this work, the author used Google AI Overview and the standard version of Grammarly to learn definitions and concepts and improve language readability, respectively. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Acknowledgments

I would like to extend my gratitude to Dr. Hamidreza Shaye and Dr. Lauren Tetz for providing me with support and guidance in writing this review. I would like to thank my mom, Dr. Arathi Shah, for inspiring me to research the potential of using CRISPR-Cas in SCT and aiding me in the creative process.

References

1. Daltveit, D. S.; Morgan, E.; Colombet, M.; Steliarova-Foucher, E.; Bendahhou, K.; Marcos-Gragera, R.; Rongshou, Z.; Smith, A.; Wei, H.; Soerjomataram, I. Global Patterns of Leukemia by Subtype, Age, and Sex in 185 Countries in 2022. *Leukemia* **2025**, *39* (2), 412–419. https://doi.org/10.1038/s41375-024-02452-y.
2. Du, M.; Chen, W.; Liu, K.; Wang, L.; Hu, Y.; Mao, Y.; Sun, X.; Luo, Y.; Shi, J.; Shao, K.; Huang, H.; Ye, D. The Global Burden of Leukemia and Its Attributable Factors in 204 Countries and Territories: Findings from the Global Burden of Disease 2019 Study and Projections to 2030. *J Oncol* **2022**, *2022*, 1612702. https://doi.org/10.1155/2022/1612702.
3. Chennamadhavuni, A.; Iyengar, V.; Mukkamalla, S. K. R.; Shimanovsky, A. Leukemia. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2026.
4. Bispo, J. A. B.; Pinheiro, P. S.; Kobetz, E. K. Epidemiology and Etiology of Leukemia and Lymphoma. *Cold Spring Harb Perspect Med* **2020**, *10* (6), a034819. https://doi.org/10.1101/cshperspect.a034819.
5. National Cancer Institute. *Acute Myeloid Leukemia - Cancer Stat Facts*. SEER. https://seer.cancer.gov/statfacts/html/amyl.html.
6. Azad, N. A.; Shah, Z. A.; Pandith, A. A.; Rasool, R.; Jeelani, S. Real-Time Quantitative PCR: A Reliable Molecular Diagnostic and Follow-up Tool for "minimal Residual Disease" Assessment in Chronic Myeloid Leukemia. *Biosci Rep* **2018**, *38* (5), BSR20180974. https://doi.org/10.1042/BSR20180974.
7. Bains, A.; Luthra, R.; Medeiros, L. J.; Zuo, Z. FLT3 and NPM1 Mutations in Myelodysplastic Syndromes: Frequency and Potential Value for Predicting Progression to Acute Myeloid Leukemia. *Am J Clin Pathol* **2011**, *135* (1), 62–69. https://doi.org/10.1309/AJCPEI9XU8PYBCIO.
8. Blanco, G.; Puiggros, A.; Baliakas, P.; Athanasiadou, A.; García-Malo, M. D.; Collado, R.; Xochelli, A.; Rodríguez-Rivera, M.; Ortega, M.; Calasanz, M. J.; Luño, E.; Vargas, M. T.; Grau, J.; Martínez-Laperche, C.; Valiente, A.; Cervera, J.; Anagnostopoulos, A.; Gimeno, E.; Abella, E.; Stalika, E.; Hernández-Rivas, J. M.; Ortuño, F. J.; Robles, D.; Ferrer, A.; Ivars, D.; González, M.; Bosch, F.; Abrisqueta, P.; Stamatopoulos, K.; Espinet, B.; Blanco, G.; Puiggros, A.; Baliakas, P.; Athanasiadou, A.; García-Malo, M. D.; Collado, R.; Xochelli, A.; Rodríguez-Rivera, M.; Ortega, M.; Calasanz, M. J.; Luño, E.; Vargas, M. T.; Grau, J.; Martínez-Laperche, C.; Valiente, A.; Cervera, J.; Anagnostopoulos, A.; Gimeno, E.; Abella, E.; Stalika, E.; Hernández-Rivas, J. M.; Ortuño, F. J.; Robles, D.; Ferrer, A.; Ivars, D.; González, M.; Bosch, F.; Abrisqueta, P.; Stamatopoulos, K.; Espinet, B. Karyotypic Complexity Rather than Chromosome 8 Abnormalities Aggravates the Outcome of Chronic Lymphocytic Leukemia Patients with TP53 Aberrations. *Oncotarget* **2016**, *7* (49), 80916–80924. https://doi.org/10.18632/ONCOTARGET.13106.
9. Boer, J. M.; Steeghs, E. M. P.; Marchante, J. R. M.; Boeree, A.; Beaudoin, J. J.; Beverloo, H. B.; Kuiper, R. P.; Escherich, G.; Van Der Velden, V. H. J.; Van Der Schoot, C. E.; De Groot-Kruseman, H. A.; Pieters, R.; Den Boer, M. L. Tyrosine Kinase Fusion Genes in Pediatric *BCR-ABL1* -like Acute Lymphoblastic Leuke-

DOI: 10.36838/IJHSR819.16

mia. *Oncotarget* **2017**, *8* (3), 4618–4628. https://doi.org/10.18632/oncotarget.13492.
10. Buccheri, V.; Barreto, W. G.; Fogliatto, L. M.; Capra, M.; Marchiani, M.; Rocha, V. Prognostic and Therapeutic Stratification in CLL: Focus on 17p Deletion and P53 Mutation. *Ann Hematol* **2018**, *97* (12), 2269–2278. https://doi.org/10.1007/s00277-018-3503-6.
11. Langabeer, S. E.; McCarron, S. L.; Kelly, J.; Krawczyk, J.; McPherson, S.; Perera, K.; Murphy, P. T. Chronic Myeloid Leukemia with E19a2 BCR-ABL1 Transcripts and Marked Thrombocytosis: The Role of Molecular Monitoring. *Case Reports in Hematology* **2012**, *2012*, 1–3. https://doi.org/10.1155/2012/458716.
12. Morris, C. M. Chronic Myeloid Leukemia: Cytogenetic Methods and Applications for Diagnosis and Treatment. In *Cancer Cytogenetics*; Campbell, L. J., Ed.; Humana Press: Totowa, NJ, **2011**; Vol. 730, pp 33–61. https://doi.org/10.1007/978-1-61779-074-4_4.
13. Schneider, F.; Hoster, E.; Schneider, S.; Dufour, A.; Benthaus, T.; Kakadia, P. M.; Bohlander, S. K.; Braess, J.; Heinecke, A.; Sauerland, M. C.; Berdel, W. E.; Buechner, T.; Woermann, B. J.; Feuring-Buske, M.; Buske, C.; Creutzig, U.; Thiede, C.; Zwaan, M. C.; Van Den Heuvel-Eibrink, M. M.; Reinhardt, D.; Hiddemann, W.; Spiekermann, K. Age-Dependent Frequencies of NPM1 Mutations and FLT3-ITD in Patients with Normal Karyotype AML (NK-AML). *Ann Hematol* **2012**, *91* (1), 9–18. https://doi.org/10.1007/s00277-011-1280-6.
14. Zainuddin, N.; Murray, F.; Kanduri, M.; Gunnarsson, R.; Smedby, K. E.; Enblad, G.; Jurlander, J.; Juliusson, G.; Rosenquist, R. TP53 Mutations Are Infrequent in Newly Diagnosed Chronic Lymphocytic Leukemia. *Leukemia Research* **2011**, *35* (2), 272–274. https://doi.org/10.1016/j.leukres.2010.08.023.
15. Zhang, L.; Habeebu, S. S. M.; Li, W. Prognostic and Predictive Biomarkers in Precursor B-Cell Acute Lymphoblastic Leukemia. In *Leukemia*; Department of Pathology and Laboratory Medicine, Children's Mercy Hospital, University of Missouri-Kansas City School of Medicine, Kansas City, USA, Weijie, L., Eds.; Exon Publications, **2022**; pp 165–193. https://doi.org/10.36255/exon-publications-leukemia-biomarkers-lymphoblastic-leukemia.
16. Tripathi, A. K.; Chuda, R. Laboratory Evaluation of Acute Leukemia. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), **2026**.
17. Choi, H.-S.; Kim, B. S.; Yoon, S.; Oh, S.-O.; Lee, D. Leukemic Stem Cells and Hematological Malignancies. *Int J Mol Sci* **2024**, *25* (12), 6639. https://doi.org/10.3390/ijms25126639.
18. De Boer, B.; Sheveleva, S.; Apelt, K.; Vellenga, E.; Mulder, A. B.; Huls, G.; Jacob Schuringa, J. The IL1-IL1RAP Axis Plays an Important Role in the Inflammatory Leukemic Niche That Favors Acute Myeloid Leukemia Proliferation over Normal Hematopoiesis. *Haematologica* **2021**, *106* (12), 3067–3078. https://doi.org/10.3324/haematol.2020.254987.
19. Hamid, G. A. Acute Leukemia Clinical Presentation. In *Leukemia*; IntechOpen, **2013**. https://doi.org/10.5772/53531.
20. Khwaja, A.; Bjorkholm, M.; Gale, R. E.; Levine, R. L.; Jordan, C. T.; Ehninger, G.; Bloomfield, C. D.; Estey, E.; Burnett, A.; Cornelissen, J. J.; Scheinberg, D. A.; Bouscary, D.; Linch, D. C. Acute Myeloid Leukaemia. *Nat Rev Dis Primers* **2016**, *2*, 16010. https://doi.org/10.1038/nrdp.2016.10.
21. Debela, D. T.; Muzazu, S. G.; Heraro, K. D.; Ndalama, M. T.; Mesele, B. W.; Haile, D. C.; Kitui, S. K.; Manyazewal, T. New Approaches and Procedures for Cancer Treatment: Current Perspectives. *SAGE Open Med* **2021**, *9*, 20503121211034366. https://doi.org/10.1177/20503121211034366.
22. Ye, W.; Wu, X.; Zhao, R.; Jin, X.; Li, H.; Qu, Y.; Ji, J.; Liu, Z. Comparison of Allo-SCT, Auto-SCT and Chemotherapy for the Treatment of Patients with Low- or Intermediate-Risk Acute Myeloid Leukemia: A Network Meta-Analysis. *Stem Cell Res Ther* **2024**, *15*, 153. https://doi.org/10.1186/s13287-024-03766-5.
23. Chevillon, F.; Dhédin, N.; Boissel, N. Allogeneic Hematopoietic Stem Cell Transplantation in Adults with Acute Lymphoblastic Leukemia: Evolving Indications and Modalities in Shifting Landscape. *Leuk Lymphoma* **2026**, *67* (2), 282–294. https://doi.org/10.1080/10428194.2025.2584685.
24. Mangum, D. S.; Caywood, E. A Clinician's Guide to HLA Matching in Allogeneic Hematopoietic Stem Cell Transplant. *Hum Immunol* **2022**, *83* (10), 687–694. https://doi.org/10.1016/j.humimm.2022.03.002.
25. Medhasi, S.; Chantratita, N. Human Leukocyte Antigen (HLA) System: Genetics and Association with Bacterial and Viral Infections. *J Immunol Res* **2022**, *2022*, 9710376. https://doi.org/10.1155/2022/9710376.
26. Montgomery, R. A.; Tatapudi, V. S.; Leffell, M. S.; Zachary, A. A. HLA in Transplantation. *Nat Rev Nephrol* **2018**, *14* (9), 558–570. https://doi.org/10.1038/s41581-018-0039-x.
27. Budeus, B.; Álvaro-Benito, M.; Crivello, P. HLA-DM and HLA-DO Interplay for the Peptide Editing of HLA Class II in Healthy Tissues and Leukemia. *Best Pract Res Clin Haematol* **2024**, *37* (2), 101561. https://doi.org/10.1016/j.beha.2024.101561.
28. Bayarsaikhan, D.; Bayarsaikhan, G.; Lee, B. Recent Advances in Stem Cells and Gene Editing: Drug Discovery and Therapeutics. *Prog Mol Biol Transl Sci* **2021**, *181*, 231–269. https://doi.org/10.1016/bs.pmbts.2021.01.019.
29. Gaj, T.; Gersbach, C. A.; Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-Based Methods for Genome Engineering. Trends *Biotechnol* **2013**, *31* (7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004.
30. Lee, J.; Sheen, J. H.; Lim, O.; Lee, Y.; Ryu, J.; Shin, D.; Kim, Y. Y.; Kim, M. Abrogation of HLA Surface Expression Using CRISPR/Cas9 Genome Editing: A Step toward Universal T Cell Therapy. *Sci Rep* **2020**, *10* (1), 17753. https://doi.org/10.1038/s41598-020-74772-9.
31. Lyu, Z.; Niu, S.; Fang, Y.; Chen, Y.; Li, Y.-R.; Yang, L. Addressing Graft-versus-Host Disease in Allogeneic Cell-Based Immunotherapy for Cancer. *Exp Hematol Oncol* **2025**, *14* (1), 66. https://doi.org/10.1186/s40164-025-00654-3.
32. Lanza, R.; Russell, D. W.; Nagy, A. Engineering Universal Cells That Evade Immune Detection. *Nat Rev Immunol* **2019**, *19* (12), 723–733. https://doi.org/10.1038/s41577-019-0200-1.
33. Dever, D. P.; Bak, R. O.; Reinisch, A.; Camarena, J.; Washington, G.; Nicolas, C. E.; Pavel-Dinu, M.; Saxena, N.; Wilkens, A. B.; Mantri, S.; Uchida, N.; Hendel, A.; Narla, A.; Majeti, R.; Weinberg, K. I.; Porteus, M. H. CRISPR/Cas9 β-Globin Gene Targeting in Human Haematopoietic Stem Cells. *Nature* **2016**, *539* (7629), 384–389. https://doi.org/10.1038/nature20134.
34. Wendt, A. S.; Brintrup, J.; Waid, J. L.; Kader, A.; Lambrecht, N. J.; Gabrysch, S. Thalassemia and Hemoglobinopathy Prevalence in a Community-Based Sample in Sylhet, Bangladesh. *Orphanet J Rare Dis* **2023**, *18* (1), 192. https://doi.org/10.1186/s13023-023-02821-3.
35. Śliwa-Tytko, P.; Kaczmarska, A.; Lejman, M.; Zawitkowska, J. Neurotoxicity Associated with Treatment of Acute Lymphoblastic Leukemia Chemotherapy and Immunotherapy. *Int J Mol Sci* **2022**, *23* (10), 5515. https://doi.org/10.3390/ijms23105515.
36. Damián-Blanco, P.; Ahuexoteco-Sánchez, S.; Carbajal-Gallardo, A. A.; Coctecon-Chávelas, F. C.; Rodríguez-Nava, C.; Vences-Velázquez, A.; Medina-Flores, Y.; Mata-Ruíz, O.; Lloret-Sánchez, L.; Cortés-Sarabia, K. Use of Monoclonal Antibodies in Cancer Immunotherapy: Types and Mechanisms of

Action. *Bol Med Hosp Infant Mex* **2023**, *80* (3), 153–164. https://doi.org/10.24875/BMHIM.23000123.
37. Bartoszewska, E.; Tota, M.; Kisielewska, M.; Skowron, I.; Sebastianka, K.; Stefaniak, O.; Molik, K.; Rubin, J.; Kraska, K.; Choromańska, A. Overcoming Antigen Escape and T-Cell Exhaustion in CAR-T Therapy for Leukemia. *Cells* **2024**, *13* (18), 1596. https://doi.org/10.3390/cells13181596.
38. Iyoda, T.; Yamasaki, S.; Ueda, S.; Shimizu, K.; Fujii, S.-I. Natural Killer T and Natural Killer Cell-Based Immunotherapy Strategies Targeting Cancer. *Biomolecules* **2023**, *13* (2), 348. https://doi.org/10.3390/biom13020348.
39. Lian, G.; Mak, T. S.-K.; Yu, X.; Lan, H.-Y. Challenges and Recent Advances in NK Cell-Targeted Immunotherapies in Solid Tumors. *Int J Mol Sci* **2021**, *23* (1), 164. https://doi.org/10.3390/ijms23010164.
40. Delgado, M.; Garcia-Sanz, J. A. Therapeutic Monoclonal Antibodies against Cancer: Present and Future. *Cells* **2023**, *12* (24), 2837. https://doi.org/10.3390/cells12242837.
41. Jo, W.; Won, T.; Daoud, A.; Čiháková, D. Immune Checkpoint Inhibitors Associated Cardiovascular Immune-Related Adverse Events. *Front Immunol* **2024**, *15*, 1340373. https://doi.org/10.3389/fimmu.2024.1340373.
42. Shyam Sunder, S.; Sharma, U. C.; Pokharel, S. Adverse Effects of Tyrosine Kinase Inhibitors in Cancer Therapy: Pathophysiology, Mechanisms and Clinical Management. *Signal Transduct Target Ther* **2023**, *8* (1), 262. https://doi.org/10.1038/s41392-023-01469-6.
43. Sugita, M.; Yamazaki, T.; Alhomoud, M.; Martinet, J.; Latouche, J.-B.; Golden, E.; Boyer, O.; Van Besien, K.; Formenti, S. C.; Galluzzi, L.; Guzman, M. L. Radiation Therapy Improves CAR T Cell Activity in Acute Lymphoblastic Leukemia. *Cell Death Dis* **2023**, *14* (5), 305. https://doi.org/10.1038/s41419-023-05829-6.
44. Halaburda, K.; Labopin, M.; Mailhol, A.; Socié, G.; Craddock, C.; Aljurf, M.; Beelen, D.; Cornelissen, J. J.; Bourhis, J.-H.; Labussière-Wallet, H.; Blaise, D.; Gedde-Dahl, T.; Gilleece, M.; Yakoub-Agha, I.; Mufti, G.; Esteve, J.; Mohty, M.; Nagler, A. Allogeneic Stem Cell Transplantation in Second Complete Remission for Core Binding Factor Acute Myeloid Leukemia: A Study from the Acute Leukemia Working Party of the European Society for Blood and Marrow Transplantation. *Haematologica* **2020**, *105* (6), 1723–1730. https://doi.org/10.3324/haematol.2019.222810.
45. Arias-Espinosa, L.; Acosta-Medina, A. A.; Vargas-España, A.; Fuentes-Martin, V.; Colunga-Pedraza, P. R.; Hawing-Zarate, J. A.; Leon, A. G.-D.; Soto-Mota, A.; Pacheco-Gutierrez, G.; Vargas-Serafín, C.; Barrera-Lumbreras, G.; Bourlon, C. Acute Leukemia Relapse after Hematopoietic Stem Cell Transplantation: The Good, the Bad, and the Ugly of Isolated Extramedullary Relapse in a Latin American Population. *Transplant Cell Ther* **2023**, *29* (8), 510.e1-510.e9. https://doi.org/10.1016/j.jtct.2023.05.006.
46. Dessie, G.; Derbew Molla, M.; Shibabaw, T.; Ayelign, B. Role of Stem-Cell Transplantation in Leukemia Treatment. *Stem Cells Cloning* **2020**, *13*, 67–77. https://doi.org/10.2147/SCCAA.S262880.
47. Boissel, N. Where Do Immunotherapies Stand in the Management of Acute Leukemia in Adults? *Hematology* **2025**, *2025* (1), 342–352. https://doi.org/10.1182/hematology.2025000723.
48. Kantarjian, H.; Jabbour, E. Adult Acute Lymphoblastic Leukemia: 2025 Update on Diagnosis, Therapy, and Monitoring. *Am J Hematol* **2025**, *100* (7), 1205–1231. https://doi.org/10.1002/ajh.27708.
49. Miller, K. D.; Nogueira, L.; Devasia, T.; Mariotto, A. B.; Yabroff, K. R.; Jemal, A.; Kramer, J.; Siegel, R. L. Cancer Treatment and Survivorship Statistics, 2022. *CA A Cancer J Clinicians* **2022**, *72* (5), 409–436. https://doi.org/10.3322/caac.21731.
50. Wagle, N. S.; Nogueira, L.; Devasia, T. P.; Mariotto, A. B.; Yabroff, K. R.; Islami, F.; Jemal, A.; Alteri, R.; Ganz, P. A.; Siegel, R. L. Cancer Treatment and Survivorship Statistics, 2025. *CA Cancer J Clin* **2025**, *75* (4), 308–340. https://doi.org/10.3322/caac.70011.
51. Champlin, R. Selection of Autologous or Allogeneic Transplantation. In *Holland-Frei Cancer Medicine. 6th edition*; BC Decker, 2003.
52. Worel, N.; Mahmoud Aljurf; Anthias, C.; Buser, A.; Cody, M.; Fechter, M.; Galeano, S.; Greinix, H.; Kisch, A. M.; Koh, M.; Thilo Mengling; Grazia Nicoloso; Dietger Niederwieser; Pulsipher, M. A.; Seber, A.; Shaw, B. E.; Stefanski, H. E.; Switzer, G. E.; Szer, J.; Walraven, van. Suitability of Hematopoietic Cell Donors: Updated Consensus Recommendations from the WBMT Standing Committee on Donor Issues. **2022**, *9* (8), e605–e614. https://doi.org/10.1016/s2352-3026(22)00184-3.
53. Deuse, T.; Hu, X.; Gravina, A.; Wang, D.; Tediashvili, G.; De, C.; Thayer, W. O.; Wahl, A.; Garcia, J. V.; Reichenspurner, H.; Davis, M. M.; Lanier, L. L.; Schrepfer, S. Hypoimmunogenic Derivatives of Induced Pluripotent Stem Cells Evade Immune Rejection in Fully Immunocompetent Allogeneic Recipients. *Nat Biotechnol* **2019**, *37* (3), 252–258. https://doi.org/10.1038/s41587-019-0016-3.
54. Kim, J.; Nam, Y.; Jeon, D.; Choi, Y.; Choi, S.; Hong, C. P.; Kim, S.; Jung, H.; Park, N.; Sohn, Y.; Rim, Y. A.; Ju, J. H. Generation of Hypoimmunogenic Universal iPS Cells through HLA-Type Gene Knockout. *Exp Mol Med* **2025**, *57* (3), 686–699. https://doi.org/10.1038/s12276-025-01422-3.
55. Uddin, F.; Rudin, C. M.; Sen, T. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. *Front Oncol* **2020**, *10*, 1387. https://doi.org/10.3389/fonc.2020.01387.
56. Daniel-Moreno, A.; Lamsfus-Calle, A.; Raju, J.; Antony, J. S.; Handgretinger, R.; Mezger, M. CRISPR/Cas9-Modified Hematopoietic Stem Cells-Present and Future Perspectives for Stem Cell Transplantation. *Bone Marrow Transplant* **2019**, *54* (12), 1940–1950. https://doi.org/10.1038/s41409-019-0510-8.
57. *A tool for genome editing: CRISPR-Cas9*. https://www.biorender.com/blog/nobel-prize-chemistry-2020 (accessed 2026-06-11).
58. Ma, X.; Chen, X.; Jin, Y.; Ge, W.; Wang, W.; Kong, L.; Ji, J.; Guo, X.; Huang, J.; Feng, X.-H.; Fu, J.; Zhu, S. Small Molecules Promote CRISPR-Cpf1-Mediated Genome Editing in Human Pluripotent Stem Cells. *Nat Commun* **2018**, *9* (1), 1303. https://doi.org/10.1038/s41467-018-03760-5.
59. De Novellis, D.; Cacace, F.; Caprioli, V.; Wierda, W. G.; Mahadeo, K. M.; Tambaro, F. P. The TKI Era in Chronic Leukemias. *Pharmaceutics* **2021**, *13* (12), 2201. https://doi.org/10.3390/pharmaceutics13122201.

Authors

Rishabh Shah is a senior at Westlake High School. He is passionate about oncogenetics and leads local stem cell drives with DKMS and NMDP to diversify the global donor registry. He hopes to major in biochemistry and the classics in the future. This review explores his interest in using gene editing tools to make stem cell transplantation safer and more accessible.

DOI: 10.36838/IJHSR819.16

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Enhancing Air Quality Forecasting by Predicting Volatile Organic Compounds Using Machine Learning

Shreyas R. Rajalingari

Monta Vista High School, 21804 McClellan Rd, Cupertino, CA, 95014, USA; go4shreyas@gmail.com

ABSTRACT: Although current research on air quality reveals that certain VOCs (like formaldehyde), which have large mass increases (>33%) during wildfire events, may be relevant for air quality prediction, they are not consistently and accurately reported or included in AQI. These issues have also severely limited the prediction of VOCs into the future. This paper attempts to test the accuracy of various machine learning models for predicting formaldehyde given historical AQI, month, and wildfire count data for counties in California from 2015 to 2025. Based on experimental results, the correlation coefficient was greater than 0.3 in a majority of cases and ranged to around 0.8, suggesting at least a moderate, positive correlation between AQI and formaldehyde. In addition, the Gradient Boosting Regressor (GBR) model performed the best (without overfitting) for predicting formaldehyde based on AQI, month, and wildfire count. The GBR model achieved an RMSE of 1.36 after hyperparameter tuning. This research suggests that there is a correlation between AQI and formaldehyde that should be explored further through higher-quality formaldehyde sensors and more consistent reporting of formaldehyde. Incorporating meteorological data and deep learning models may improve the accuracy of formaldehyde prediction as well.

KEYWORDS: Earth and Environmental Science, Atmospheric Science, Air Quality, Volatile Organic Compounds, Formaldehyde.

Introduction

The Air Quality Index (AQI) is the most widely adopted system for reporting air quality across the United States. The Environmental Protection Agency currently calculates AQI by selecting the highest pollutant index among 5 regulated criteria pollutants: particulate matter (less than 2.5 and 10 microns), ozone, carbon monoxide, sulfur dioxide, and nitrogen dioxide. The pollutant index is defined as follows:

$$\text{Pollutant Index} = \frac{AQI_{Hi} - AQI_{Low}}{Conc_{Hi} - Conc_{Low}} \times (Conc_i - Conc_{Low}) + AQI_{Low}$$

In the equation above, $Conc_i$ refers to the measured concentration of the pollutant, $Conc_{Hi}$ and $Conc_{Low}$ refer to the pollutant-specific breakpoints immediately greater than and less than the measured concentration, respectively, and AQI_{Hi} and AQI_{Low} refer to the corresponding AQI value for the Hi and Low breakpoints, respectively. The pollutant-specific breakpoints come from the EPA National Ambient Air Quality Standards (NAAQS).[1]

Despite its widespread usage, the current AQI system has several limitations. For example, the criteria pollutants used for AQI do not include larger particles that are produced during wildfires, including ash. In addition, these pollutants do not include volatile organic compounds (VOCs), which are harmful toxins often produced by wildfires. These omissions may be especially important in states with historically high wildfire counts, such as California. One study by Jaffe *et al.* notes that in addition to particulate matter, wildfire smoke can contain VOCs, including benzene, toluene, xylene, acrolein, and formaldehyde,[2] which are associated with certain health risks like cancer and neurological defects. Compared to the 5 criteria pollutants measured in AQI, VOCs are often unreliably reported due to limited monitoring frequency and poor-quality sensors, making it difficult to obtain an accurate model for air quality that integrates them.

Another study by Rice *et al.* shows that formaldehyde increased by around 34% on average during wildfires, the second largest mean mass increase during wildfire events among all VOCs (second to acrolein at 43%).[3] These results show that certain VOCs may be meaningful predictors of general air quality. Since formaldehyde was also more consistently tracked by the EPA than acrolein (although still to an irregular extent), this paper will focus on formaldehyde as a case study.

Existing ML-based air quality forecasting studies are typically limited to AQI or the 5 criteria pollutants that are found in AQI. For example, a study conducted by Fan *et al.* forecasted ozone and $PM_{2.5}$ in the Pacific Northwest,[4] while a study by Raffuse *et al.* forecasted $PM_{2.5}$ in wildfire conditions.[5] However, neither of these studies considers volatile organic compounds. This highlights a clear gap in research focusing on the forecasting of VOCs like formaldehyde, which can have equally consequential health effects.

Current research on air quality reveals that while certain VOCs are relevant for air quality prediction and have dangerous health effects, they aren't accurately reported or included in AQI. This paper attempts to examine the relationship between AQI and formaldehyde and assess whether AQI actually predicts formaldehyde concentration in order to measure the efficacy of extrapolating AQI to the concentrations of harmful VOCs. In addition, while substantial research exists attempting to predict AQI into the future, the prediction of formaldehyde has been severely limited due to the inadequate monitoring of formaldehyde and the lower quality of formaldehyde sensors.

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.17

This paper attempts to close this gap by testing the accuracy of various machine learning models for predicting formaldehyde given historical AQI, month, and wildfire count information for counties in California from 2015 to 2025. Data from California was specifically chosen to more accurately determine the relevance of wildfires in formaldehyde prediction, as California is known to be a wildfire hotspot. This paper first attempts to establish a correlation between AQI and formaldehyde and then trains various simple machine learning models on AQI, month, and fire count (the number of wildfires in the county of the measurement on that particular date) and attempts to accurately predict formaldehyde. It was hypothesized that there is a positive correlation between AQI and formaldehyde, with AQI, month, and fire count all being important factors for predicting formaldehyde.

Methods

Data:

The AQI and VOC datasets[6] are publicly available and are found on the EPA website, specifically under Pre-generated Data Files on the EPA Air Quality System (AQS) page. The specific files used are Hourly VOC data and Daily AQI data from 2015 through 2025. The temporal range 2015 through 2025 was chosen to obtain the most recent and accurate VOC data. Entries for VOC data are typically averages taken over 1, 3, 8, or 24 hours of a single day, while all entries for AQI data are averages taken over the entire day. The wildfire data[7] is also publicly available and found on the CALFire website under Incidents/Incident Data.

Materials:

Google Colaboratory, a hosted Jupyter Notebook service, was used as the development environment and is where all of the code for this paper was written. All of the code for this paper was written in Python. The data downloaded from the sources listed in the Data subsection above was stored in a folder on Google Drive, and Google Colaboratory allowed this drive to be mounted directly to the notebook itself. The following Python libraries were used during code development: pandas, sklearn, pathlib, typing, matplotlib, glob, and os. The pandas, pathlib, typing, glob, and os libraries were used to compile the data and build the pandas DataFrames, while sklearn was used to build the models themselves.

Data Preprocessing:

1. AQI Data Preprocessing: First, the CSV file containing the AQI dataset was converted to a pandas DataFrame. Then it was filtered by state name such that only California AQI data from 2015 to 2025 remained.

2. VOC Data Preprocessing: First, the CSV file containing the VOC dataset was converted to a pandas DataFrame. Then the DataFrame was filtered by state name and parameter name such that only formaldehyde data from California from 2015 to 2025 remained. In cases where the formaldehyde data contained multiple entries for the same date and location, the entries were deduplicated by keeping rows in the following order of precedence: highest sample duration (time over which the sample was taken), then highest arithmetic mean (concentration of formaldehyde for that entry). Higher sample durations were prioritized because they provide estimates that are more accurate for an entire day. The highest arithmetic mean was prioritized because the maximum concentration was also the most dangerous concentration for that location on that date.

3. Merging the Datasets: First, the AQI and VOC datasets were merged using the columns for county name and date to keep only entries where both AQI and VOC data were captured. Then the CALFire data was used to add a column called "fire counts," which counted the number of wildfires burning in that entry's location for that particular day.

4. AQI and VOC data were compared to establish any correlation between them. This step is described in more detail in the Data Analysis subsection.

5. Removing Outliers and Compiling the Final Dataset: First, data where the VOC observation percent—the percentage of samples taken out of the total scheduled to be taken—was less than 50% were removed to remove data that was primarily incomplete. This step also removed all but four of the outlier entries for the formaldehyde data because most of the original outliers had observation percentages that were less than 50%. Since the number of remaining outliers was extremely small compared to the number of observations, they had a negligible impact on model performance. Finally, the dataset was filtered to 6 columns, which comprised the final dataset that was used to train the machine learning models: state name, county name, date, AQI, formaldehyde, and fire count. Training the machine learning models is described in more detail in the Models subsection.

Data Analysis:

1. Normalizing the dataset: First, all of the columns in a copy of the final dataset were z-score normalized using StandardScaler from the sklearn library.

2. Establishing a Relationship Between AQI and Formaldehyde: To identify a relationship between AQI and formaldehyde, a dual-axis line plot was created for every entry in the dataset with AQI and normalized formaldehyde plotted on the y-axis (with an equal scale for both) and the date plotted on the x-axis. The correlation coefficient between normalized AQI and normalized formaldehyde was then calculated to quantify the correlation between AQI and formaldehyde.

Models:

1. Splitting Feature and Target Datasets: The feature DataFrame, containing the AQI, month (came from date), and fire count columns, was separated from the target DataFrame containing the formaldehyde column.

2. Normalizing the Feature Dataset: All of the features in the feature DataFrame were z-score normalized using a StandardScaler.

3. Splitting the Data into Testing and Training Datasets: Both the feature and target DataFrames were separated into training and testing data, using the train_test_split function

from sklearn, such that 75% of the data was used for training and 25% was used for testing.

4. Building and Testing the Baseline Models: First, a Linear Regressor was built using the open-source model found in the sklearn library. The linear regression model attempts to find the line of best fit for data that is linearly correlated. The Linear Regressor was trained on the training set and tested using the testing dataset. The following regression line was found after testing:

$$formaldehyde = 3.0504 + 0.9221 \times AQI + 0.1709 \times Fire\ Count + 0.2814 \times Month$$

A dummy regressor from the sklearn library was also built in order to provide a baseline to measure how much each model improved over a useless model. The dummy regressor always predicts the average of the target variable.

5. Building and Testing the Models: First, the following models were built from open-source models found in the sklearn library: Random Forest Regressor, Gradient Boosting Regressor, XGBoost, LightGBM, and SVR. The Random Forest Regressor trains many individual decision trees and then averages the results. The Gradient Boosting Regressor sequences many decision trees, with each one learning from the errors of the previous model. The XGBoost and LightGBM models are variants of the Gradient Boosting Regressor. The SVR model maps data into high-dimensional space and attempts to find the best-fitting hyperplane to minimize loss. Next, each of the models was trained on the training set and tested using the testing dataset.

6. Tuning the Models: After training and testing, the hyperparameters of each model (where applicable) were tuned using GridSearchCV from the sklearn library. GridSearchCV searches through all combinations of a specified hyperparameter dictionary to find the combination with the best average R-squared score (percentage of variance in the target variable captured by the input features). Finally, the root mean squared error (RMSE) metric was used to assess the performance of each model, and the best-performing model was identified by selecting the one that improved the most over the baseline after hyperparameter tuning. Percentage improvement over the baseline was used as the ranking metric to highlight the relative usefulness of each machine learning model over a naive approach, even though raw RMSE after tuning would yield the same ranking.

Results and Discussion

Results:

First, dual-axis line plots were created showing normalized AQI and formaldehyde vs date for each of the counties in the time period from 2015 to 2025 in order to establish a relationship between AQI and formaldehyde. Each line plot also plotted semi-translucent orange bands in the background, which corresponded to the periods when each fire for that county was active. After each of the plots was plotted, the correlation coefficient between AQI and formaldehyde was calculated for each. Three examples of these plots are shown below.

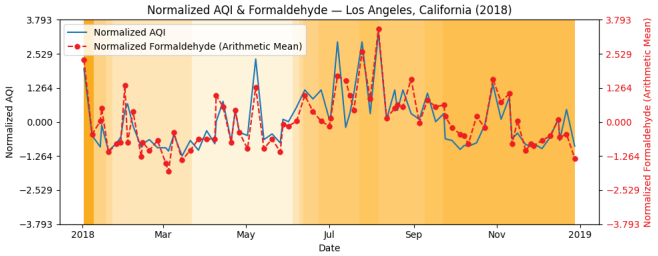


Figure 1: Dual-axis line plot of normalized AQI and formaldehyde vs date for Los Angeles County in 2018, before removing outliers, with semi-translucent orange bands plotted in the background to represent the periods during which each fire for 2018 in Los Angeles County was active. The correlation coefficient between normalized AQI and normalized Formaldehyde is approximately 0.8012.

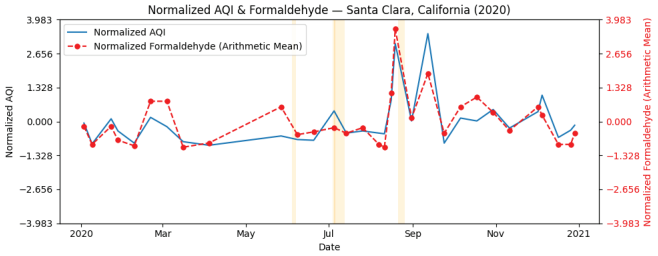


Figure 2: Dual-axis line plot of normalized AQI and formaldehyde vs date for Santa Clara County in 2020, before removing outliers, with semi-translucent orange bands plotted in the background to represent the periods during which each fire for 2020 in Santa Clara County was active. The correlation coefficient between normalized AQI and normalized Formaldehyde is approximately 0.8459.

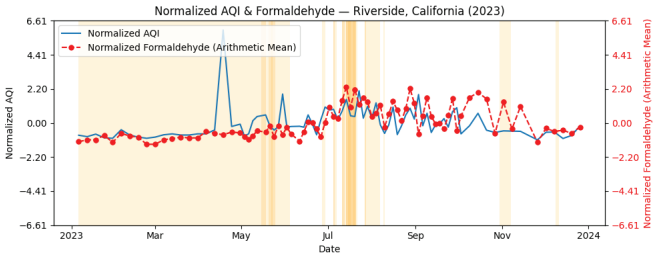


Figure 3: Dual-axis line plot of normalized AQI and formaldehyde vs date for Riverside County in 2023, before removing outliers, with semi-translucent orange bands plotted in the background to represent the periods during which each fire for 2023 in Riverside County was active. The correlation coefficient between normalized AQI and normalized Formaldehyde is approximately 0.3450.

After a relationship between AQI and formaldehyde was established, the Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, XGBoost, LightGBM, and SVR models were built and tested. Each of the models was trained on the training set and tested on the testing dataset. After this, each of the models, with the exception of the Linear Regression, was hyperparameter-tuned using GridSearchCV. After this, a dummy regressor was built in to provide a baseline, and the improvement of the hyperparameter-tuned models over the dummy regressor was measured as a percentage. The results are summarized below.

DOI: 10.36838/IJHSR819.17

Table 1: RMSE on the testing dataset before and after hyperparameter tuning, the percentage of RMSE improvement over the Dummy Regressor, and the best hyperparameters found after tuning using GridSearchCV for each model. The Gradient Boosting Regression and XGBoost models had the highest percent RMSE improvement over the dummy regressor.

Model	RMSE Prior to Hyperparameter Tuning	RMSE After Hyperparameter Tuning	Percent RMSE Improved Over Dummy Regressor	Best Hyperparameters Found using GridSearchCV
Dummy Regressor	1.98	N/A	N/A	N/A
Linear Regression	1.88	N/A	5.05	N/A
Random Forest Regression	1.57	1.39	29.8	max_depth: 10, max_features: "sqrt", min_samples_split: 10, n_estimators: 100
Gradient Boosting Regression	1.38	1.36	31.31	learning_rate: 0.1, max_depth: 3, n_estimators: 100
XGBoost	1.56	1.36	31.31	colsample_bytree: 0.7, learning_rate: 0.05, max_depth: 3, n_estimators: 200, subsample: 0.9
LightGBM	1.40	1.37	30.81	learning_rate: 0.5, max_depth: 3, n_estimators: 100, num_leaves: 20, reg_alpha: 0.5, reg_lambda: 0.1
SVR	1.40	1.38	30.30	C: 1, epsilon: 0.2, kernel: "rbf"

The percentage increase in RMSE from training to testing for each of the models was also calculated in order to establish which models overfit and underfit the data. These results are summarized below.

Table 2: RMSE on the training and testing dataset after hyperparameter tuning using GridSearchCV (when applicable) and percentage increase in RMSE for each model. The XGBoost model had a higher percentage increase in RMSE from training to testing than the Gradient Boosting Regression model.

Model	RMSE on the Training Dataset	RMSE on the Testing Dataset	Percentage Increase in RMSE From Training to Testing
Dummy Regressor	1.90	1.98	4.04
Linear Regression	1.59	1.88	15.43
Random Forest Regression	1.17	1.39	15.83
Gradient Boosting Regression	1.27	1.36	6.62
XGBoost	0.97	1.36	28.68
LightGBM	1.20	1.37	12.41
SVR	1.36	1.38	1.45

Table 3: Mean Absolute Error (MAE) and R^2 on the testing dataset after hyperparameter tuning using GridSearchCV for each model.

Model	Mean Absolute Error (MAE) on the Testing Dataset After Hyperparameter Tuning	R^2 on the Testing Dataset After Hyperparameter Tuning
Random Forest Regression	1.070	0.367
Gradient Boosting Regression	0.922	0.510
XGBoost	1.031	0.375
LightGBM	0.954	0.495
SVR	0.917	0.498

Tree-based ensemble models like Gradient Boosting Regressor are highly interpretable, allowing for feature importance to be assessed. Below is the feature importance for each of the features used to train the tuned Gradient Boosting Regressor model.

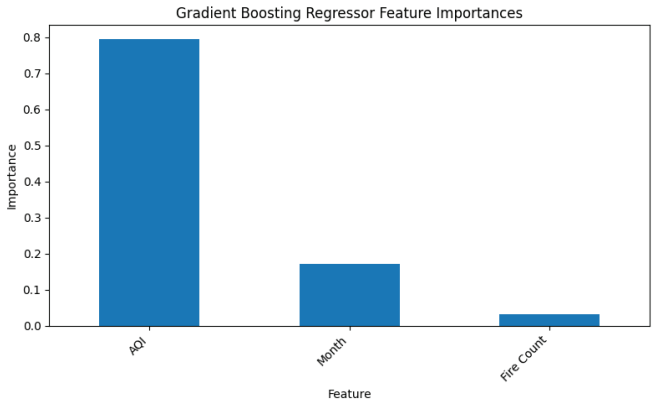


Figure 4: Bar chart showing the importance of each feature for the Gradient Boosting Regressor model after hyperparameter tuning. AQI was the most important, the month was the second most important, and fire count was the least important.

Finally, the Gradient Boosting Model predictions were plotted on a scatterplot showing the true values for formaldehyde for each normalized AQI value in the testing dataset. The results are shown below.

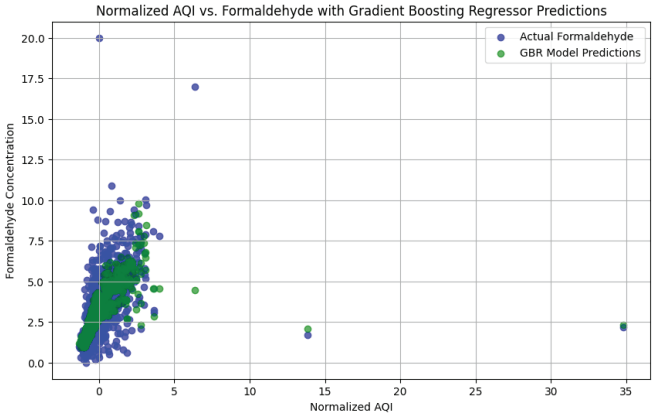


Figure 5: Scatter plot showing formaldehyde vs normalized AQI for each entry in the testing dataset and the Gradient Boosting Regressor model's formaldehyde predictions in green for each entry in the testing dataset. The Gradient Boosting Regressor model's predictions were close to the actual formaldehyde values.

Discussion:

A majority of the calculated correlation coefficients between AQI and formaldehyde were greater than 0.3, indicating that there was at least a moderate positive correlation between AQI and formaldehyde. In some cases, for counties where a greater amount of formaldehyde data existed, the correlation coefficient was greater, suggesting a stronger correlation. For example, for Los Angeles County in 2018 (Figure 1), a county that had a high level of formaldehyde that year had a correlation coefficient of approximately 0.8012, which indicated a strong positive correlation. However, this was not always the case, as seen in Riverside County in 2023 (Figure 3), which had a correlation coefficient of approximately 0.3450 despite high levels of formaldehyde reporting. Outliers in the AQI dataset, especially the large spike in mid- to late April, may have contributed to this weaker relationship. This conclusion is supported by the fact that qualitatively there was a clear correlation between AQI and formaldehyde. In addition, it can

also be determined qualitatively that AQI and formaldehyde generally had greater values when there were many fires than when there were no fires.

Based on the results in Table 1, it can be established that the Gradient Boosting Regression and XGBoost models performed best on the testing dataset since those models had the greatest improvement in RMSE (31.31%) over the dummy regressor, which was the baseline model. Gradient Boosting Regression and XGBoost are models that perform well for complex, non-linear relationships between the features and the target, which is exactly what was found during the data analysis from Figures 1-3. However, after hyperparameter tuning, each of the models, except for linear regression, performed similarly overall, suggesting that if careful hyperparameter tuning is performed, any non-linear model could be a viable one.

Based on the results in Table 2, it is likely that XGBoost overfitted the training dataset, since its percentage increase in RMSE was relatively high, which suggests overfitting. The Gradient Boosting Regressor model, however, did not seem to overfit the data because its percentage increase in RMSE was relatively low. This suggests that the Gradient Boosting Regressor model was able to generalize better, making it the best overall model to predict formaldehyde based on the AQI, month, and fire count. This conclusion is supported by Table 3, which shows that the Gradient Boosting Regressor model had the second-lowest MAE and the highest R^2 value.

From the feature importances in Figure 4, it can be seen that AQI was the most important feature at around 0.79 importance, and fire count was the least important at around 0.04. This shows that AQI was an important feature in predicting formaldehyde, which furthers the idea that AQI is a good predictor of formaldehyde overall. The fire count was the least important, which was surprising given that formaldehyde tends to increase during fires. The dominance of AQI suggests that it tracks short-term VOC fluctuations well, while fire activity contributes only indirectly due to its primarily long-term effects.

Based on Figure 5, it can be seen that the Gradient Boosting Regression model's formaldehyde predictions are relatively close to the actual values in the testing data and that the predictions seem to capture the overall trend between AQI and formaldehyde well.

Limitations:

While it can preliminarily be established that there is at least a moderate positive correlation between AQI and formaldehyde, more consistent reporting and better formaldehyde data quality are necessary to determine a more precise correlation. Also, AQI is likely an incomplete predictor because the same AQI value can have different compositions of the 5 criteria pollutants, which this experiment did not account for.

Additionally, it is important to note that there are confounding factors like meteorological conditions and quality differences between AQI sensors and VOC monitors that should be examined further to establish a more definitive correlation or relationship. Such factors, or noise in the data, may explain the four outliers in Figure 5, where formaldehyde spiked without a corresponding increase in AQI or AQI spiked without a corresponding increase in formaldehyde.

When training each of the models, the amount of data was only around 5000 rows, and there may still have been noise in the data (particularly in the AQI data), both of which could lead to overfitting. If the amount of high-quality data were increased, it is likely that the issue of overfitting seen in most of the models would subside, but further analysis with more data needs to be conducted to test that conjecture.

Furthermore, the initial data split did not preserve temporal ordering, which may have caused data leakage during training because the model could learn from future data to predict past observations. Any future repetitions of this experiment should preserve temporal ordering to alleviate this risk. However, when tested, preserving temporal ordering yielded negligible differences in results.

Next Steps:

One step that can be taken to potentially improve the models is to incorporate meteorological data as features. This could include wind speed, carbon monoxide, humidity, and temperature, which Khan *et al.* found to be the most important features for predicting AQI.[8] This suggests that these features could be important for predicting formaldehyde as well, due to the established relationship between AQI and formaldehyde. In addition, training deep learning models or neural networks should be tested as the next step to evaluate whether those models can achieve better performance than the simpler models that were used for this paper.

■ Conclusion

This paper examined the relationship between AQI, month, wildfires, and formaldehyde by first establishing a relationship between AQI and formaldehyde and then training 6 regression models to see how accurately AQI, month, and fire count can predict formaldehyde using historical data from California. From the experiments conducted, it can be concluded that there is at least a moderate, positive correlation between AQI and formaldehyde. In addition, the Gradient Boosting Regressor was the best of the models tested for predicting formaldehyde based on AQI, month, and fire count. When tuned properly, however, all of the non-linear models produced similar results, suggesting that, with proper tuning, any of the non-linear models tested can be viable for predicting formaldehyde. Surprisingly, fire count was not as important a feature as the other two for predicting formaldehyde. With more high-quality data, a more precise relationship between AQI and formaldehyde can be calculated. Finally, incorporating meteorological data or deep learning models should be tested as a next step as a potential way to improve model performance.

■ Acknowledgments

I would like to acknowledge Shiang Chin, PhD. for his invaluable guidance throughout the project.

DOI: 10.36838/IJHSR819.17

References

1. *AQI Breakpoints | Air Quality System | US EPA*. aqs.epa.gov. https://aqs.epa.gov/aqsweb/documents/codetables/aqi_breakpoints.html.
2. Jaffe, D. A.; O'Neill, S. M.; Larkin, N. K.; Holder, A. L.; Peterson, D. L.; Halofsky, J. E.; Rappold, A. G. Wildfire and Prescribed Burning Impacts on Air Quality in the United States. *Journal of the Air & Waste Management Association* **2020**, *70* (6). https://doi.org/10.1080/10962247.2020.1749731.
3. R. Byron Rice; Boaggio, K.; Olson, N. E.; Foley, K. M.; Weaver, C. P.; Sacks, J. D.; McDow, S. R.; Holder, A. L.; LeDuc, S. D. Wildfires Increase Concentrations of Hazardous Air Pollutants in Downwind Communities. *Environmental Science & Technology* **2023**. https://doi.org/10.1021/acs.est.3c04153.
4. Fan, K.; Dhammapala, R.; Harrington, K.; Lamb, B.; Lee, Y. Machine learning-based ozone and $PM_{2.5}$ forecasting: Application to multiple AQS sites in the Pacific Northwest. *Frontiers* **2023**. https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2023.1124148/full.
5. Raffuse, S.; O'Neill, S.; Schmidt, R. *A model for rapid $PM_{2.5}$ exposure estimates in wildfire conditions using routinely available data: Rapidfire v0.1.3*. Geoscientific Model Development 2024. https://doi.org/10.5194/gmd-17-381-2024.
6. *AirData website File Download page*. aqs.epa.gov. https://aqs.epa.gov/aqsweb/airdata/download_files.html#Annual.
7. CAL FIRE. *2025 Fire Season Outlook | Welcome to CAL FIRE*. Ca.gov. https://www.fire.ca.gov/incidents.
8. Khan, R. R.; Dwivedi, S.; Yadav, S.; Sharf, R.; Maurya, S.; Pandey, U. P.; Kumari, S. Integrating AI with IoT Sensors for Real-Time Air Quality Monitoring and Prediction. *International Journal of Environmental Sciences* **2025**, 2274–2287. https://doi.org/10.64252/pc0vbk56.

Author

Shreyas Rajalingari is a student at Monta Vista High School in Cupertino, CA. His intended major in college is computer science, AI, or data science. He loves to play football and baseball, hike, and bike. He also likes to travel around the world and meet new people.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Applications of CRISPR/Cas9 in Modeling and Investigating Schizophrenia

Siya Gupta

Saratoga High School, 20300 Herriman Ave, Saratoga, CA, 95070, USA; 14siya.universe@gmail.com

ABSTRACT: Schizophrenia is an impairing psychiatric disorder that causes patients to experience distorted perceptions of reality, altering their regular thoughts, emotions, actions, and feelings. This disease currently lingers as an unknown condition whose precise biological mechanisms and genetic links remain incompletely understood due to a lack of research, and therefore has no decisive cure. This lack of knowledge has intensified the need for biomedical tools that are capable of examining schizophrenia at a genetic level. CRISPR/Cas9 is a 21st-century genome-editing technique that offers the potential to investigate the cellular processes underlying many under-researched diseases. This tool can accurately cut DNA at specific locations, enabling scientists to replicate or modify targeted genetic regions. In this context, CRISPR/Cas9 provides a formidable approach for studying complex, polygenic disorders such as schizophrenia. Using CRISPR/Cas9, scientists can develop biomedical technologies, such as animal models, epigenome-editing tools, and induced pluripotent stem cells, that can model specific genetic variants implicated in schizophrenia. These tools will magnify and isolate these mutations, providing scientists with a clearer view of the intricacies of this mental disorder. However, this genome-editing tool is still in the research stage and has recently been discovered, making it essential to carefully consider its scientific challenges, limitations, and ethical implications.

KEYWORDS: Medical and Health Sciences, Psychiatric Diseases, Schizophrenia, Genetics, Polygenic Risk Variants, CRISPR/Cas9 Genome Editing.

Introduction

Schizophrenia is a crippling mental disorder that affects 1 percent of the world's population.[1] This illness affects an individual's thoughts, emotions, actions, and feelings by altering their perception of reality.[2] This psychiatric disease is polygenic, meaning that thousands of genetic mutations are suspected of contributing to the risk. As of now, schizophrenia remains a challenging mental disorder; a definitive cure that erases damage on a genetic level has yet to be discovered. While current treatment options can help manage symptoms and improve quality of life, they cannot remove the root cause of this disease, which lies in the patient's genes. The large number and complexity of the genetic factors that bring about schizophrenia make it so difficult for researchers to study; each patient can be diagnosed with a different combination of mutations, triggering different effects. However, an emerging genome-editing tool, CRISPR/Cas9, offers an opportunity to correct and replicate genetic mutations associated with schizophrenia. Studies in animal models and cell culture systems are already providing valuable insights into how CRISPR/Cas9 can be applied safely and effectively. By precisely targeting and modifying genes, researchers could use CRISPR/Cas9 to develop mouse models, epigenetic editing tools, and induced pluripotent stem cells (iPSCs) that incorporate genetic variants identified in patients with schizophrenia, thereby providing valuable insights into disease mechanisms.[3] These CRISPR/Cas9-based mechanisms will allow researchers to isolate specific linked mutations and test their effects on neural function on a larger scale. With a better understanding of schizophrenia, genome editing could potentially revolutionize the way schizophrenia and other complex diseases are treated, shifting the focus from symptom control to true cure and prevention in the future. Although this emerging technology is still grappling with ethical concerns and scientific limitations, increased research and funding in the future could further develop this tool into a highly effective method for studying and potentially treating schizophrenia. In this review, I will provide an overview of the CRISPR/Cas9 genome-editing tool, followed by a discussion of its applications for modeling and studying schizophrenia. Finally, I will examine the therapeutic potential of CRISPR/Cas9 in schizophrenia, along with the current limitations and challenges associated with its use.

The overview of schizophrenia:

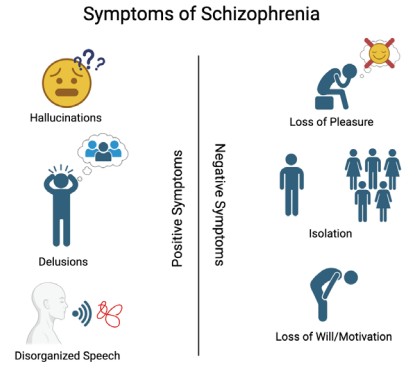


Figure 1: Symptoms of Schizophrenia. The figure above depicts common symptoms that are indicative of schizophrenia. The left side classifies positive symptoms: hallucinations, delusions, and disorganized speech; and the right side classifies negative symptoms: loss of pleasure, isolation, and loss of will/motivation, of schizophrenia. Created in Biorender.

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.18

Schizophrenia is characterized by positive and negative symptoms that can influence a patient's thoughts, perceptions, speech, affect, and behaviors, as depicted in Figure 1.[2] Positive symptoms include hallucinations, particularly auditory hallucinations involving voices that communicate with or about the patient, and paranoid delusions. Negative symptoms include reduced emotional expression, loss of pleasure, loss of will and motivation, and social withdrawal.[4] Schizophrenia is also characterized by disorganized thoughts, which are highlighted in speech and behavior. Disorganized speech may range from loose associations and rapid movement among multiple topics to speech so muddled that it resembles schizophrenia (commonly referred to as "word salad") (add any paper?). This disorganized behavior that schizophrenia causes may lead to difficulties in performing daily living activities, such as preparing meals or maintaining personal hygiene. It can also manifest as childlike silliness or outbursts of unpredictable anger and irritation. Schizophrenia is also associated with death rates that are two to three times higher than those seen in the general population. This surplus in deaths is accounted for by a combination of an increased risk of suicide, unhealthy lifestyles, and inadequate healthcare that schizophrenia patients go through.[5] Many individuals with schizophrenia experience mood-related symptoms such as depression and anxiety, which may contribute to the 10% lifetime incidence of suicide in schizophrenia.[4]

Mutation	Main Function	Relation to Schizophrenia
DISC1	helps regulate brain development, neuron growth, and neuron connection	can disrupt communication between neurons involved in memory, emotion, and reasoning
FMR1	controls synapse development and protein production at synapses	changes in the gene may affect neuron communication linked to schizophrenia
NRG1	involved in brain development, myelination, and neuron communication	can cause abnormal brain wiring and disrupted neuron signaling
3q29 deletion	contains genes involved in mitochondrial and neuron function	linked to mitochondrial inflexibilty and increased schizophrenia risk due to malfunctioning neurons
22q11.2 deletion	contains multiple genes important for dopamine regulation, neuron migration, and synaptic communication	disrupts many aspects of brain function, greatly increasing the risk of schizophrenia
CASKIN1	sets up signaling complexes at synapses	may disrupt communication between neurons
15q11.2 microdeletion	affects genes involved in synapse regulation and neuronal signaling	can increase the risk of schizophrenia by affecting synapse development and brain signaling

Figure 2: Table of Genetic Mutations Associated with Schizophrenia. The table above lists some genes and chromosomal deletions that are associated with schizophrenia. It explains how each gene functions normally in the brain, along with the reason why variations in these genes contribute to schizophrenia. Created in Google Sheets.

Schizophrenia is a complex mental disorder with no single known cause. However, among the various contributing factors - environmental, genetic, and psychosocial - genetics appears to be the most significant one. An individual's likelihood of developing schizophrenia is six times higher if they have a close relative with the genetic variations of this disease.[2] A combination of genetic mutations called Single Nucleotide Polymorphisms (SNPs) contributes to the development of schizophrenia. Some of these SNPs are outlined in Figure 2. For example, Callicott *et al.* studied 12 genetic variants in Disrupted-in-Schizophrenia 1 (DISC1), which plays a crucial role in brain development.[6] Some DISC1 mutations can disrupt the growth and connectivity of neurons in the hippocampus and prefrontal cortex, which play crucial roles in memory, emotion, and reasoning.[11] The study found that a common variant named Ser704Cys in DISC1 was found in patients with schizophrenia more often.[6] The FMR1 gene works closely with the DISC1 gene to control how neurons connect and communicate. A disruption in this gene can influence the development of schizophrenia, as many patients have an altered amount of FMRP (a protein produced by FMR1).[7] Another candidate gene is Neuregulin 1 (NRG1), which encodes a protein involved in multiple brain processes, including brain development, myelination (insulating layers around nerve fibers that speed up electrical signal transmission), and neuronal communication.[8] Certain variants in the NRG1 gene have resulted in changes in brain development, causing abnormal neural wiring and disrupted neuron communication and myelination, which are distinct characteristics of schizophrenia. Problems with some neurotransmitters, such as dopamine and glutamate, which allow neurons to communicate with each other, may also contribute to the pathophysiology of schizophrenia.[2] The 3q29 deletion is another strong indicator of an increased risk in many psychiatric disorders, including schizophrenia.[9] Specifically, researchers Purcell *et al.* found associated dysregulation of mitochondrial function in mouse models. Tests showed that cells lacked mitochondrial flexibility, meaning they were unable to adapt their energy use when conditions changed.[10] Additionally, patients with a small deleted region on chromosome 22, called the 22q11.2 deletion, have 30% increased risk of developing schizophrenia, as it gets rid of many genes important for brain development. The genes in this deleted region help break down dopamine in the brain, regulate microRNAs (which control gene expression), play important roles in neuronal communication, and influence how neurons assemble during brain development.[11] A protein called CASKIN1 helps set up signaling pathways at synapses (the connection points between neurons). A mutation in the gene that codes for this protein can pose a great risk to brain function and can aid in the development of schizophrenia.[12] Lastly, a small deletion on chromosome 15 called the 15q11.2 microdeletion is closely related to schizophrenia as it erases genes that aid in brain development and neuron communication.[13] Overall, the role that these mutations play in increasing schizophrenia risk further solidifies that this psychiatric disease is polygenic.

As of now, there is no definitive cure for schizophrenia, but it can be treated and managed in a few ways, such as antipsychotic medications and psychotherapy. Psychotherapy is a form of treatment for emotional, behavioral, and mental problems in which a trained therapist works with a patient to help them understand themselves better and develop healthy coping strategies.[14] One type of psychotherapy that has been proven to be effective for schizophrenia is Cognitive-Behavioral Therapy (CBT). CBT for schizophrenia is adapted to address psychotic symptoms, especially delusions, hallucinations, and thought disorders, and to help patients develop insight into their condition, reduce distress from symptoms, and improve coping skills.[15] Antipsychotics are medications primarily used

to manage psychosis, especially in conditions including schizophrenia and bipolar disorder. They are used to reduce positive symptoms such as hallucinations and delusions and help stabilize mood. Since dopamine overactivity is linked to psychotic symptoms, antipsychotics work by blocking dopamine D2 receptors in the brain.[16] However, despite some of these available treatment options, schizophrenia remains a condition that can not be fully cured. These therapies primarily focus on controlling symptoms and improving daily functioning rather than studying and targeting the root genetic factors responsible for the disease. The exact cause of schizophrenia is still not understood, involving a complex mix of genetics, which makes developing a definitive cure challenging as of now.[6]

The overview of the CRISPR/Cas9 system:

CRISPR/Cas9 consists of a DNA array and Cas proteins that cut DNA at specific sites. The CRISPR/Cas9 array consists of short, palindromic repeats separated by unique spacer sequences derived from foreign DNA.7 Different regions of the CRISPR/Cas9 array are copied into RNA, which eventually become CRISPR/Cas9 RNA (cRNA) and trans-activating CRISPR/Cas9 RNA(tracrRNA). These two molecules guide the CRISPR/Cas9 system to cut specific DNA sequences.[8] The crRNA is a complementary strand to the spacer retrieved from foreign DNA stored in the CRISPR/Cas9 array. The crRNA, tracrRNA, and Cas9 counter the DNA of attacking viruses. The guide RNA (gRNA) is made up of crRNA and tracrRNA.[8]

CRISPR/Cas9 is an immune system found in many bacteria and archaea.[7] When a bacterium is infected by a bacteriophage, the CRISPR/Cas9 system cuts a segment of viral DNA and stores it in the CRISPR/Cas9 array as a new spacer. Each time the same bacteriophage reinfects, the bacterium can destroy it using its guide RNA and Cas9. The CRISPR/Cas9 system compares the virus's DNA with all of its spacers. If the DNA matches one of the spacers, the bacterium can identify the invader as a threat. The guide RNA guides Cas9 to the bacteriophage, where it cleaves the DNA on both strands, creating a double-strand break (DSB). DSBs eliminate the virus because it lacks its own DNA repair mechanisms.[9]

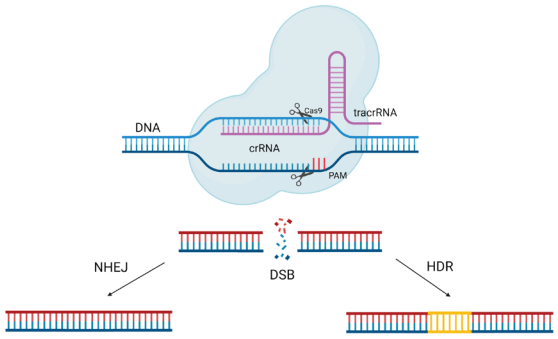


Figure 3: CRISPR/Cas9 DNA Cutting and Repair Pathways. Cas9, guided by crRNA and tracrRNA, binds to a target DNA sequence next to a PAM site and creates a double-strand break (DSB). The DNA is then repaired either by non-homologous end joining (NHEJ), which joins the broken ends together, or by homology-directed repair (HDR), which allows the deleted DNA to be replicated. Created in Biorender.

The CRISPR/Cas9 system, originally discovered as a bacterial immune response mechanism, has been adapted into a genome-editing tool to treat genetic diseases. To apply the CRISPR/Cas9 system in genetic engineering, scientists artificially modify the CRISPR's guide RNA to match the targeted cell in an organism's genome.[10] The guide RNA directs Cas9 to the location of the mutation in the organism's DNA, where it searches for the protospacer adjacent motif (PAM), a short sequence downstream of the target site.[7] When it recognizes PAM, Cas9 checks the region above. If it locates the target provided by the gRNA, it will create a DSB. The DNA will then replace the missing mutated sequence with the correct sequence, using the second strand to make a copy. DSBs are repaired by either homologous direct repair (HDR) or non-homologous end joining (NHEJ).[8] NHEJ is simple and random, while HDR is much more accurate and complex.[9] HDR uses identical donor DNA to repair DNA damage, and NHEJ joins broken DNA ends, often resulting in many missing or added DNA segments.[11]

Gene therapy has long shown potential to correct a variety of human diseases and defects. The discovery of CRISPR/Cas9 and its use as a gene-editing tool have revolutionized the field of gene therapy. Much of this enthusiasm stems from the accuracy and specificity with which CRISPR/Cas9 can treat human disease and edit the human genome.[7]

Application of CRISPR/Cas-9 in Schizophrenia Research:

Schizophrenia is a polygenic disorder, meaning hundreds of genetic variants contribute to the risk together but have small impact sizes individually.[18] The sheer number and size of these mutations hinder researchers' ability to examine their effects and learn more about this obscure disease. The CRISPR/Cas9 system offers a solution to this problem, enabling precise and accurate editing of genetic sequences through multiple approaches, including animal models, induced pluripotent stem cells, and epigenome-editing tools.[3] With the help of these genetically engineered tools, researchers can segregate specific genes that are suspected to increase schizophrenia risk and study their implications individually.

Mice are commonly used to model the human genome because their genes are highly similar to those of humans, and their brain structure closely resembles that of humans. To replicate a schizophrenia-associated mutation in a mouse, scientists use the CRISPR/Cas9 genome editing tool.

They insert a Cas9 protein, a guide RNA, and a repair template into a fertilized mouse egg. Cas9 cuts the DNA to produce the genetic variant, and the cell repairs it using the repair template. The edited egg is implanted into a surrogate female mouse to be carried to term. The first generation may contain a mixture of edited and unedited cells, so they are often bred together to produce stable experimental outcomes. As these mice develop, scientists carefully examine the changes in brain structure, protein production, gene expression, and behavior.[19] In 2019, Baba *et al.* used CRISPR/Cas9 to delete the 3q29 region in mice to replicate the schizophrenia-associated 3q29 deletion.[20] These edited mice showed 50% lower expression of most genes in the 3q29 region. They also exhibited a

DOI: 10.36838/IJHSR819.18

significant reduction in body weight, consistent with growth deficiencies observed in humans with the same deletion. Social interaction, brain function, and reaction were also skewed within these mice.[3] With these engineered mouse models, researchers were able to isolate the 3p29 deletion and pinpoint its exact effects. By observing changes in behavior and function, they were able to study the direct connection between the deletion and schizophrenia symptoms. In 2020, Saito *et al.* conducted the same experiment using a mouse model with the 22q11.2 deletion.[21] With this deletion, the mice had trouble filtering out unnecessary sensory information, remembering cues in fear-based tests, and processing their vision. These behavioral changes are characteristic of individuals who have schizophrenia.[3] Again, the researchers were able to narrow their focus on the 22q11.2 deletion and directly observe its contribution to schizophrenia on a larger scale. Mouse models like these give researchers a better understanding of schizophrenia, helping them test how specific mutations contribute to this disorder in a controlled environment. These studies demonstrate that CRISPR/Cas9-engineered mouse models can effectively replicate human genetic mutations associated with schizophrenia, thereby providing valuable tools for studying the effects of specific genes.

Epigenetics is the study of how changes in gene expression occur without altering the DNA sequence itself. These changes often involve chemical marks on DNA, such as DNA methylation, which can turn off a gene, or histone modifications, which can make DNA easier or harder for the cell to process.[22] Epigenome editing enables precise targeting of these switches to turn specific genes on or off. Instead of using traditional CRISPR/Cas9, scientists employ a dead CRISPR/Cas9 tool, dCas9, for epigenetic editing. dCas9 can still be guided to a specific DNA sequence by its gRNA, but it can not cut a nucleotide. Instead, it is fused to an epigenetic effector protein that alters the chemical marks on the target DNA. For example, dCas9 could be fused to an activator, which turns a gene on; a repressor, which turns a gene off; or an epigenetic enzyme, which fine-tunes genetic activity. Since schizophrenia is linked to SNPs that cause genes to be turned on or off at the wrong times rather than to broken genes, scientists could use epigenome editing to study how a gene's activity affects neurons and to fix abnormal gene activity in lab-grown cells, which could lead to therapies in the future. In 2018, Liu *et al.* conducted an experiment using CRISPR/Cas9 to remove the chemical tags on the FMR1 gene.[23] The chemical tags on this gene silence it, preventing it from making the protein the brain needs to function properly. The researchers used the Tet1 enzyme, which can remove these chemical tags, and aimed it directly at the silenced FMR1 gene. Soon, the chemical tags on these genes were removed, and it began producing its protein again. [24] This study showed that epigenetic editing tools can successfully be used to reverse errors in gene regulation and restore normal neural function. By comparing neurons before and after editing, researchers can study how disrupted gene expression contributes to schizophrenia symptoms. This evidence also suggests that CRISPR/Cas9 epigenetic tools can one day be used to correct irregular gene activity contributing to schizophrenia on a much larger scale.

Induced pluripotent stem cells (iPSCs) are lab-created stem cells derived from skin or blood adult cells by reprogramming them into pluripotent cells (cells that can specialize into any type of cell in the body).[26] Scientists can derive iPSCs from individuals with or without schizophrenia. Using CRISPR/Cas9, they can precisely modify a single base pair in DNA, either by targeting a specific SNP or by introducing a new SNP. This approach generates two cell populations: one with the variant and one without, both of which differentiate into neurons.[27] Scientists observe the differences in neuron behavior, and because everything else in the DNA is the same, any differences can be attributed specifically to that single genetic variant.[18] In 2023, Wahbeh *et al.* used CRISPR/Cas9 to introduce a mutation in the CASKIN1 gene (suspected of contributing to schizophrenia) into iPSCs, which were then differentiated into neurons. When they compared the neuron with the mutation with the control, they found that the expression of 368 other genes had changed significantly. These mutant neurons fired electrical signals at rates different from those of normal neurons, affecting their development over time.[3] These findings uncover the substantial effects that a single mutation can have on otherwise healthy neurons. It solidifies our understanding of schizophrenia as a genetic disorder, largely influenced by errors in DNA sequences disrupting brain function. Additionally, in 2014, Yoon *et al.* generated iPSCs from individuals with the 15q11.2 microdeletion, a known genetic variation associated with schizophrenia.[28] The 15q11.2 microdeletion contains a missing copy of a protein called CYFIP1, which controls the cytoskeleton. This is especially important in neurons, which have precise shapes and connections. Scientists observed that, in the absence of the CYFIP1 protein, patients' cells failed to adhere properly and were disorganized, resulting in many neurons being mislocalized.[3] Again, with investigations such as these, researchers are able to deepen their comprehension of how specific mutations contribute to schizophrenia. These genetic variations need to be individually dissected to obtain a better grasp on the mechanisms of this psychiatric disorder and the changes that need to be introduced to explore better treatment options in the future.

These studies show that using CRISPR/Cas9 genome editing in iPSCs can elucidate the effects of specific schizophrenia-related genetic mutations.

Therapeutic Potential and Challenges:

The CRISPR/Cas9 gene-editing tool can be directly applied to human patients for therapeutic purposes. With its ability to precisely modify a mutated DNA sequence, CRISPR/Cas9 has the potential to eliminate inherited genetic disorders right at the source. It also has the potential to combat viral infections with a higher success rate by targeting and cleaving viral genetic material within human cells. With further research, CRISPR/Cas9 can be leveraged to precisely correct human DNA, thereby enabling the development of effective treatments for many incurable genetic disorders.[29]

As promising as animal models, iPSCs, and epigenome editing seem, scientists cannot overlook the potential limitations and challenges these tools may entail. Self-evidently, animal models can only represent the complex behaviors of psychiatric disorders such as schizophrenia in a limited way. The modeled neural responses of these cells and animal models are not guaranteed to replicate the responses observed in patients with schizophrenia exactly.[30] Another challenge involves finding a safe way to deliver the Cas9 protein into cells. Researchers typically use modified viruses, such as adeno-associated viruses, baculoviral, and lentiviral vectors, to deliver the editing tools to target cells.[24] However, the body's immune system may recognize these viruses as threats, triggering an immune response that rejects the treatment.[31] The potential for the complete treatment of psychiatric diseases like schizophrenia that CRISPR/Cas9 invites is still quite far from development. This advancement is relatively new and remains in the early stages of research; it has been tested primarily on laboratory animals.[32] And because psychiatric diseases are influenced by many genes, it would be almost impossible for CRISPR/Cas9 to edit every single one of these genes at the stage it is right now.[33] Researchers also have concerns about the CRISPR/Cas9 editing tool, including mosaicism, off-target effects, and incomplete edits. Mosaicism can occur when not all mutated cells are edited the same way, resulting in different genetic profiles within the same individual.[34] Off-target effects occur when CRISPR/Cas9 accidentally cuts or edits the DNA at the wrong location. These unintended edits can disrupt essential genes, potentially leading to adverse effects.[35] And the edits implemented on the target can be incomplete or low-quality.[36]

In addition to the scientific challenges posed by the CRISPR/Cas9 genome-editing tool, there are important ethical and moral concerns regarding its use in human patients. A major point of debate among researchers concerns the ethics of germline alterations, which involve making genetic changes in an embryo's DNA that can be passed down to future generations.[36] These changes could cause unexpected and unwanted mutations in subsequent generations, as researchers still do not fully understand the associated risks. According to Lanphier *et al.*, embryo editing could result in mosaic embryos, in which not all cells share the same genetic modifications.[37] In 2015, scientists in China began using CRISPR/Cas9 on non-viable embryos under strict human research regulations, such as the Regulation on Human Genetic Resources for Experimental Purposes.[38] However, in 2019, Dr. Jiankui He went against these regulations and edited the genes of viable embryos to disable the CCR5 gene, which HIV uses to infect cells. These edited embryos were implanted despite concerns about potential additional health risks.[39] This event sparked major controversy. Overall, we still do not fully understand how the CRISPR/Cas9 gene-editing tool might affect complex diseases, such as schizophrenia, making ethical considerations even more crucial.

Conclusion

Schizophrenia remains a complex psychiatric disorder with thousands of genetic links and no definite cure. This review highlights how the CRISPR/Cas9 gene-editing tool provides a powerful new approach to studying the genetic factors that contribute to schizophrenia. Specifically, researchers can use CRISPR/Cas9 in animal models, induced pluripotent stem cells, and epigenome editing to closely examine how specific genetic mutations affect brain development and function. Although we still need to consider major limitations, such as off-target effects, immune reactions, and the polygenic nature of schizophrenia, ongoing advances in this tool suggest a promising future for its research contributions. The future direction of this developing technology should include improving the precision and safety to ensure its accuracy and reliability. Specifically for the CRISPR-based mouse models, epigenetic editing, and iPSCs, a larger number of genetic mutations should be tested to map out the contributors to schizophrenia. Overall, CRISPR/Cas9 holds the potential to advance schizophrenia research, improve our understanding of its intricate genetic connections, and pave the path for future therapeutic breakthroughs.

Acknowledgments

I would like to thank my mentor, Daniel Kim, for his guidance throughout the review process.

References

1. Insel, T. R. Rethinking Schizophrenia. *Nature* **2010**, *468* (7321), 187–193. https://doi.org/10.1038/nature09552.
2. Schultz, S. H.; North, S. W.; Shields, C. G. Schizophrenia: A Review. *Am. Fam. Physician* **2007**, *75* (12), 1821–1829.
3. Kurishev, A. O.; Karpov, D. S.; Nadolinskaia, N. I.; Goncharenko, A. V.; Golimbet, V. E. CRISPR/Cas-Based Approaches to Study Schizophrenia and Other Neurodevelopmental Disorders. Int. J. Mol. Sci. 2022, 24 (1), 241. https://doi.org/10.3390/ijms24010241.
4. Lewis, D. A.; Lieberman, J. A. Catching up on Schizophrenia: Natural History and Neurobiology. *Neuron* **2000**, *28* (2), 325–334. https://doi.org/10.1016/s0896-6273(00)00111-2.
5. Auquier, P.; Lançon, C.; Rouillon, F.; Lader, M.; Holmes, C. Mortality in Schizophrenia. *Pharmacoepidemiol. Drug Saf.* **2006**, *15* (12), 873–879. https://doi.org/10.1002/pds.1325.
6. Kantrowitz, J. T.; Correll, C. U.; Jain, R.; Cutler, A. J. New Developments in the Treatment of Schizophrenia: An Expert Roundtable. *Int. J. Neuropsychopharmacol.* **2023**, *26* (5), 322–330. https://doi.org/10.1093/ijnp/pyad011.
7. Lino, C. A.; Harper, J. C.; Carney, J. P.; Timlin, J. A. Delivering CRISPR: A Review of the Challenges and Approaches. *Drug Deliv.* **2018**, *25* (1), 1234–1257. https://doi.org/10.1080/10717544.2018.1474964.
8. Sandhya, D.; Jogam, P.; Allini, V. R.; Abbagani, S.; Alok, A. The Present and Potential Future Methods for Delivering CRISPR/Cas9 Components in Plants. *J. Genet. Eng. Biotechnol.* **2020**, *18* (1), 25. https://doi.org/10.1186/s43141-020-00036-8.
9. Shen, S.; Loh, T. J.; Shen, H.; Zheng, X.; Shen, H. CRISPR as a Strong Gene Editing Tool. *BMB Rep.* **2017**, *50* (1), 20–24. https://doi.org/10.5483/bmbrep.2017.50.1.128.
10. Asmamaw, M.; Zawdie, B. Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing. *Biol. Targets Ther.* **2021**, *15*, 353–361. https://doi.org/10.2147/BTT.S326422.
11. Miyaoka, Y.; Berman, J. R.; Cooper, S. B.; Mayerl, S. J.; Chan, A. H.; Zhang, B.; Karlin-Neumann, G. A.; Conklin, B. R. Systematic Quantification of HDR and NHEJ Reveals Effects of Locus, Nu-

DOI: 10.36838/IJHSR819.18

clease, and Cell Type on Genome-Editing. *Sci. Rep*. **2016**, *6*, 23549. https://doi.org/10.1038/srep23549.
12. Callicott, J. H.; Straub, R. E.; Pezawas, L.; Egan, M. F.; Mattay, V. S.; Hariri, A. R.; Verchinski, B. A.; Meyer-Lindenberg, A.; Balkissoon, R.; Kolachana, B.; Goldberg, T. E.; Weinberger, D. R. Variation in DISC1 Affects Hippocampal Structure and Function and Increases Risk for Schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (24), 8627–8632. https://doi.org/10.1073/pnas.0500515102.
13. Mackie, S.; Millar, J. K.; Porteous, D. J. Role of DISC1 in Neural Development and Schizophrenia. *Curr. Opin. Neurobiol*. **2007**, *17* (1), 95–102. https://doi.org/10.1016/j.conb.2007.01.007.
14. Mostaid, M. S.; Lloyd, D.; Liberg, B.; Sundram, S.; Pereira, A.; Pantelis, C.; Karl, T.; Weickert, C. S.; Everall, I. P.; Bousman, C. A. Neuregulin-1 and Schizophrenia in the Genome-Wide Association Study Era. *Neurosci. Biobehav. Rev.* **2016**, *68*, 387–409. https://doi.org/10.1016/j.neubiorev.2016.06.001.
15. Goldfried, M. R. What Should We Expect from Psychotherapy? *Clin. Psychol. Rev.* **2013**, *33* (5), 654–662. https://doi.org/10.1016/j.cpr.2012.09.006.
16. Bustillo, J.; Lauriello, J.; Horan, W.; Keith, S. The Psychosocial Treatment of Schizophrenia: An Update. *Am. J. Psychiatry* **2001**, *158* (2), 163–175. https://doi.org/10.1176/appi.ajp.158.2.163.
17. Tandon, R. Antipsychotics in the Treatment of Schizophrenia: An Overview. *J. Clin. Psychiatry* **2011**, *72 Suppl 1*, 4–8. https://doi.org/10.4088/JCP.10075su1.01.
18. Michael Deans, P. J.; Brennand, K. J. Applying Stem Cells and CRISPR Engineering to Uncover the Etiology of Schizophrenia. *Curr. Opin. Neurobiol.* **2021**, *69*, 193–201. https://doi.org/10.1016/j.conb.2021.04.003.
19. St Clair, D.; Johnstone, M. Using Mouse Transgenic and Human Stem Cell Technologies to Model Genetic Mutations Associated with Schizophrenia and Autism. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2018**, *373* (1742), 20170037. https://doi.org/10.1098/rstb.2017.0037.
20. Rutkowski, T. P.; Purcell, R. H.; Pollak, R. M.; Grewenow, S. M.; Gafford, G. M.; Malone, T.; Khan, U. A.; Schroeder, J. P.; Epstein, M. P.; Bassell, G. J.; Warren, S. T.; Weinshenker, D.; Caspary, T.; Mulle, J. G. Behavioral Changes and Growth Deficits in a CRISPR Engineered Mouse Model of the Schizophrenia-Associated 3q29 Deletion. *Mol. Psychiatry* **2021**, *26* (3), 772–783. https://doi.org/10.1038/s41380-019-0413-5.
21. Saito, R.; Koebis, M.; Nagai, T.; Shimizu, K.; Liao, J.; Wulaer, B.; Sugaya, Y.; Nagahama, K.; Uesaka, N.; Kushima, I.; Mori, D.; Maruyama, K.; Nakao, K.; Kurihara, H.; Yamada, K.; Kano, M.; Fukada, Y.; Ozaki, N.; Aiba, A. Comprehensive Analysis of a Novel Mouse Model of the 22q11.2 Deletion Syndrome: A Model with the Most Common 3.0-Mb Deletion at the Human 22q11.2 Locus. *Transl. Psychiatry* **2020**, *10* (1), 35. https://doi.org/10.1038/s41398-020-0723-z.
22. Avitzour, M.; Mor-Shaked, H.; Yanovsky-Dagan, S.; Aharoni, S.; Altarescu, G.; Renbaum, P.; Eldar-Geva, T.; Schonberger, O.; Levy-Lahad, E.; Epsztejn-Litman, S.; Eiges, R. FMR1 Epigenetic Silencing Commonly Occurs in Undifferentiated Fragile X-Affected Embryonic Stem Cells. *Stem Cell Rep*. **2014**, *3* (5), 699–706. https://doi.org/10.1016/j.stemcr.2014.09.001.
23. Liu, X. S.; Wu, H.; Krzisch, M.; Wu, X.; Graef, J.; Muffat, J.; Hnisz, D.; Li, C. H.; Yuan, B.; Xu, C.; Li, Y.; Vershkov, D.; Cacace, A.; Young, R. A.; Jaenisch, R. Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. *Cell* **2018**, *172* (5), 979-992.e6. https://doi.org/10.1016/j.cell.2018.01.012.
24. Paquet, D.; Kwart, D.; Chen, A.; Sproul, A.; Jacob, S.; Teo, S.; Olsen, K. M.; Gregg, A.; Noggle, S.; Tessier-Lavigne, M. Efficient Introduction of Specific Homozygous and Heterozygous Mutations Using CRISPR/Cas9. *Nature* **2016**, *533* (7601), 125–129. https://doi.org/10.1038/nature17664.
25. Powell, S. K.; Gregory, J.; Akbarian, S.; Brennand, K. J. Application of CRISPR/Cas9 to the Study of Brain Development and Neuropsychiatric Disease. *Mol. Cell. Neurosci*. **2017**, *82*, 157–166. https://doi.org/10.1016/j.mcn.2017.05.007.
26. Robinton, D. A.; Daley, G. Q. The Promise of Induced Pluripotent Stem Cells in Research and Therapy. *Nature* **2012**, *481* (7381), 295–305. https://doi.org/10.1038/nature10761.
27. Gutiérrez-Rodríguez, A.; Cruz-Fuentes, C. S.; Genis-Mendoza, A. D.; Nicolini, H. CRISPR/Cas9 Genome Editing Approaches for Psychiatric Research. *Rev. Bras. Psiquiatr*. **2023**, *45* (2), 137–145. https://doi.org/10.47626/1516-4446-2022-2913.
28. Yoon, K.-J.; Nguyen, H. N.; Ursini, G.; Zhang, F.; Kim, N.-S.; Wen, Z.; Makri, G.; Nauen, D.; Shin, J. H.; Park, Y.; Chung, R.; Pekle, E.; Zhang, C.; Towe, M.; Hussaini, S. M. Q.; Lee, Y.; Rujescu, D.; St Clair, D.; Kleinman, J. E.; Hyde, T. M.; Krauss, G.; Christian, K. M.; Rapoport, J. L.; Weinberger, D. R.; Song, H.; Ming, G.-L. Modeling a Genetic Risk for Schizophrenia in iPSCs and Mice Reveals Neural Stem Cell Deficits Associated with Adherens Junctions and Polarity. *Cell Stem Cell* **2014**, *15* (1), 79–91. https://doi.org/10.1016/j.stem.2014.05.003.
29. Piergentili, R.; Del Rio, A.; Signore, F.; Umani Ronchi, F.; Marinelli, E.; Zaami, S. CRISPR-Cas and Its Wide-Ranging Applications: From Human Genome Editing to Environmental Implications, Technical Limitations, Hazards and Bioethical Issues. *Cells* **2021**, *10* (5), 969. https://doi.org/10.3390/cells10050969.
30. LaMarca, E. A.; Powell, S. K.; Akbarian, S.; Brennand, K. J. Modeling Neuropsychiatric and Neurodegenerative Diseases With Induced Pluripotent Stem Cells. *Front. Pediatr.* **2018**, *6*, 82. https://doi.org/10.3389/fped.2018.00082.
31. Memi, F.; Ntokou, A.; Papangeli, I. CRISPR/Cas9 Gene-Editing: Research Technologies, Clinical Applications and Ethical Considerations. *Semin. Perinatal.* **2018**, *42* (8), 487–500. https://doi.org/10.1053/j.semperi.2018.09.003.
32. Zhu, H.; Zhang, L.; Tong, S.; Lee, C. M.; Deshmukh, H.; Bao, G. Spatial Control of *in Vivo* CRISPR-Cas9 Genome Editing via Nanomagnets. *Nat. Biomed. Eng*. **2019**, *3* (2), 126–136. https://doi.org/10.1038/s41551-018-0318-7.
33. Sakai, C.; Ijaz, S.; Hoffman, E. J. Zebrafish Models of Neurodevelopmental Disorders: Past, Present, and Future. *Front. Mol. Neurosci.* **2018**, *11*, 294. https://doi.org/10.3389/fnmol.2018.00294.
34. Hsu, P. D.; Lander, E. S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *157* (6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010.
35. Ran, F. A.; Hsu, P. D.; Wright, J.; Agarwala, V.; Scott, D. A.; Zhang, F. Genome Engineering Using the CRISPR-Cas9 System. *Nat. Protoc*. **2013**, *8* (11), 2281–2308. https://doi.org/10.1038/nprot.2013.143.
36. Brokowski, C.; Adli, M. CRISPR Ethics: Moral Considerations for Applications of a Powerful Tool. *J. Mol. Biol.* **2019**, *431* (1), 88–101. https://doi.org/10.1016/j.jmb.2018.05.044.
37. Lanphier, E.; Urnov, F.; Haecker, S. E.; Werner, M.; Smolenski, J. Don't Edit the Human Germ Line. *Nature* **2015**, *519* (7544), 410–411. https://doi.org/10.1038/519410a.
38. Zhai, X.; Ng, V.; Lie, R. No Ethical Divide between China and the West in Human Embryo Research. *Dev. World Bioeth*. **2016**, *16* (2), 116–120. https://doi.org/10.1111/dewb.12108.
39. Cyranoski, D. Russian Biologist Plans More CRISPR-Edited Babies. *Nature* **2019**, *570* (7760), 145–146. https://doi.org/10.1038/d41586-019-01770-x.

Author

Siya Gupta is a high school sophomore who lives in the Bay Area and attends Saratoga High. She has a deep interest and passion for the education of biomedical engineering and will take any opportunity to learn more about the subject. She decided to write this paper for this exact reason; she wanted to address a crucial real-world problem while broadening her knowledge at the same time. In her free time, Siya enjoys swimming, playing guitar, and reading.

DOI: 10.36838/IJHSR819.18

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

A Functional and Socioeconomic Analysis of State-of-the-Art Prostheses Selection: Comparing Traditional, Myoelectric, and AI-Powered Devices

Sheraya M. Amarasekara

Dubai College, Al Sufouh - Al Sufouh 2 - Dubai, Dubai, 837, UAE; sma.78r@gmail.com

ABSTRACT: Worldwide, an estimated 35–40 million people require prosthetic or orthotic devices. Prosthetic devices are categorized as traditional, myoelectric, or AI-powered. Traditional prostheses are most widely used due to affordability and accessibility, but offer limited functional complexity. Myoelectric prostheses address some limitations by utilizing electromyographic signals from the residual limb to enable more complex movements; however, they require substantial training and remain constrained by needing manual mode switching. AI-powered prostheses build on myoelectric systems by applying machine learning to interpret user intent more accurately, facilitating automatic mode switching and refined control; however, they carry greater financial and weight-related barriers. Although each type offers advantages and drawbacks, selection ultimately depends on user preference, social norms, and clinical feasibility. Current literature tends to focus on individual device types, lacking a holistic, multi-factor comparison to guide informed decision-making. This study addresses that gap by analyzing and comparing literature across five factors: (i) Control and Comfort; (ii) Adaptability; (iii) Cost and Accessibility; (iv) Emotional Benefits and Social Preference; and (v) Durability, Robustness, and Simplicity. Overall, AI-powered devices provide the most advanced control and adaptability, traditional prostheses remain the most durable and accessible, and myoelectric devices occupy an intermediate position. This paper provides a centralized resource supporting evidence-based prosthesis selection.

KEYWORDS: Biomedical Engineering, Biomedical Devices, Prosthetics, Myoelectric, Artificial Intelligence.

Introduction

A prosthetic device is any artificial body part that replaces the function or appearance of missing anatomy.[1] These devices are widely used, with over 35–40 million people globally requiring prosthetic or orthotic services. Prosthetic devices can be broadly divided into two types: traditional and digital. A traditional prosthesis is usually a mechanical device that uses physical power from the residual limb to facilitate movement, replacing a body part that is missing or no longer functional.[2] They could be either passive aesthetic restorations or use a harness and cables for tension-based control.[3] They stand within a well-established market and consist of a range of fairly accessible products. However, due to their antiquated nature and reliance on organic sources of force, such as the sound limb, for example,[4] they have certain shortcomings. These include limited function, given the lack of electronic components, and lower comfort levels.[5] To mitigate these shortcomings, more developed digital options are arising within the world of biomedical research.

Digital prostheses can be divided into two types: myoelectric prostheses or AI-powered ones. Myoelectric prostheses are externally powered and operate by interpreting electromyographic (EMG) signals from muscles in the residual limb through pairs of electrodes and amplifiers to replicate natural joint movement.[6] The user trains their muscles to make certain movements/contractions that correspond to actions the prosthetic makes, such as grip changes, opening or closing the hand, etc.[7] They require less organic force from the user, and thus less energy, reducing muscle strain.[8]

Myoelectric prostheses use electrodes to capture EMG signals, which can only control one degree of freedom (e.g., opening and closing the hand).[9] To control other functions, mode switching allows users to change grip types by accurately producing triggers with their residual limb,[10] which can activate other degrees of freedom.[11] This can provide a greater range of functionality with a limited number of electrodes, but requires extensive training to employ the technique.

On the other hand, AI-powered digital prostheses refer to any electrical/mechanical device that uses artificial intelligence to interpret gathered data on the movements of a user, producing more natural prosthetic movement. They often use recurrent neural networks to allow real-time decoding of nerve signals for intuitive control, with an accuracy of up to 98% for individual finger and wrist movements of myoelectric hands.[12] In contrast to myoelectric prostheses, AI-powered devices learn from the user's EMG signal data (strength, location, etc.) to ensure that the prosthesis can interpret the user's intended movement more accurately, thus executing the action with higher precision.[13] Digital prostheses mitigate the shortcomings of traditional devices by generating their own force for movement, having a wider range of more natural, human-like movement, and learning from the user's movements to provide a more comfortable experience.[14] However, they lead to additional constraints related to cost, comfort, and accessibility.[15]

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.19

Both upper and lower limb prostheses are discussed similarly in the paper, but distinctions are made where needed. Particularly in the analysis of digital devices, control discussions are mainly confined to the upper limb where more precise control is required (in the fingers in comparison with the toes, for example).

Digital solutions are transforming prosthetic care, yet replacing traditional prostheses with their digital counterparts presents limitations. Most of the aforementioned research is often device-specific; those studies that do offer comparison often focus on functional differences rather than comprehensively reviewing all factors related to use. This paper addresses this gap by evaluating traditional, myoelectric, and AI-powered prostheses across five factors: (i) Control and Comfort; (ii) Adaptability; (iii) Cost and Accessibility; (iv) Emotional Benefits and Social Preference; and (v) Durability, Robustness, and Simplicity. The aim is to provide a holistic comparison of all device types, which can be further developed into a centralized repository for prosthesis selection, allowing clinicians and users to make more informed choices. The findings will also help identify future research directions where trade-offs, such as cost and weight, can be mitigated.

Factors Affecting Prosthesis Choice

Literature was searched to assess the factors affecting prosthesis choice. Inclusion criteria were as follows: papers published within the 1990-2026 period, papers published in English, and papers directly referencing prostheses, myoelectric prostheses, or AI-powered prostheses. Exclusion criteria were papers older than 1990 and those written in other languages.

Control and Comfort:

Prosthetic control includes the methods through which a patient can manipulate their prosthetic device for use in daily living. Traditional prosthetic devices are typically passive, but can use body-powered harnesses for control.[16] While simple to do, these movements can be conspicuous and unnatural.[17]

In contrast, myoelectric prosthetic control is neuromuscular and uses EMG signals. Surface electrodes are placed over the residual limb muscles;[18] patients contract these muscles, and the electrodes detect the electrical nerve impulses produced from this movement, called EMG signals. These EMG signals are used to generate the action to be carried out within the prosthesis.[19] Further, incorporating AI into prosthetics allows machine-learning algorithms to interpret and learn from user data, such as EMG signals, to provide a full range of motion, easier mode switching, and detailed variation in grip strength,[13,20] thus improving control. For example, the random forest algorithm has been used to classify additional types of motion beyond the two co-contraction movements typically required to 'trigger' actions in myoelectric prostheses, allowing higher precision.[21,22]

While control is essential for enabling precise, reliable movement, the comfort of a prosthesis is critical to ensure continued use.[23] Prosthetic comfort indicates the physical ease or lack of pain experienced during device operation.

All three prosthetic types use similar custom-fit plastic sockets to interface between the body and device. However, traditional sockets can be uncomfortable if they rely on body-powered harnesses, which require high input force from the user.[24] Myoelectric sockets have surface electrodes that can feel uncomfortable against the skin and have added weight. AI-powered sockets have the same structure as myoelectric ones, but with added weight due to additional sensors, circuits, and batteries. This often makes digital prostheses heavier than traditional ones.[25]

Overall, digital devices offer more advanced control compared to traditional ones by using EMG signaling. Further, by improving mode switching using machine-learning algorithms, AI-powered prostheses enable higher precision than myoelectric prostheses.

Adaptability:

Prosthetic adaptability refers to the ability to modify the device's components and functions to suit the user's needs. High adaptability enables the prosthesis to feel like an extension of the body and maximizes functionality.

Traditional prostheses often struggle to mimic natural motion and may function only in binary states (e.g., prosthetic hands often rely on a limited open/close function). Therefore, for different activities, terminal devices (Figure 1) must be used, which form the end part of a prosthesis that interacts with the patient's environment. Their functions vary from activity-specific prostheses (such as for biking or push-ups) to passive restorations that serve only aesthetic purposes.[26] While they enhance functionality, frequent removal and replacement can be tedious and expensive.

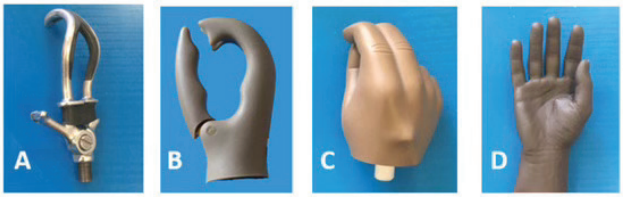


Figure 1: Four examples of terminal devices. (A) Split hook, (B) Adept Prehensor, (C) CAPP hand, (D) Passive cosmesis. Reproduced from ref 26 under Creative Commons License CC BY 4.0. Copyright 2023 The Authors. Published by MDPI.[27]

Myoelectric prostheses can allow users to perform varied tasks (producing different grip types, opening or closing the hand, etc.) through mode switching. Users will generate 'triggers' by flexing different muscles depending on which action they want to execute.[10, 28] Furthermore, AI-driven prostheses use machine learning (ML) to continuously learn and adapt to a user's movements over time.[29] This facilitates more personalized and intuitive control, leading to a more natural feel and improved functionality.[15]

Traditional prostheses typically require minimal training to use.30 In contrast, myoelectric control is unintuitive to most and can require extensive training and calibration; AI-based prostheses can overcome this issue by allowing signals to be interpreted more easily over time and progressively 'learning' user behavior, enabling more intuitive, fine-grained control.[14] Nonetheless, there is still a learning curve involved in the

training of muscles in the residual limb to control a limb that is no longer there.

Overall, to vary function, traditional prostheses rely on interchangeable terminal devices, while myoelectric prostheses use mode switching. AI-powered prostheses use pattern recognition algorithms to better interpret the user's EMG signals over time, improving the precision of mode switching.

Cost and Accessibility:

The cost of a prosthetic device is related to several factors, including materials, custom fabrication and fitting, technological complexity, ongoing research and development (R&D), as well as clinical services and follow-up care required for proper fitting, calibration, and training. These high costs can not only deter patients from prosthetic devices but also create barriers to access. Since prosthetic use requires specialized services, which are not available at all hospitals, accessibility is limited by capacity constraints at hospitals and proximity to clinics with appropriate resources.[31] A hospital visit to get a prosthetic device prescribed is a financial burden in and of itself.[32]

A major cost associated with prostheses is direct attachment to the human body using a personalized prosthetic socket. The level of personalization required for every prosthetic arm makes high costs inevitable.[33] This raises ethical concerns, especially as those with socio-economic weakness are found to have higher rates of lower-limb amputation, a finding that can be extrapolated to the upper-limb cohort.[34]

Traditional prostheses, while less expensive than their digital counterparts, still can cost from $3000 to $10,000, with those that have more advanced mechanical or hydraulic systems being priced at $15,000.[35] However, global insurance coverage for prostheses remains limited, with many limbs deemed 'medically unnecessary,'[36] despite their established role in healthcare. This has led to only 30% of beneficiaries from 2016 to 2019 receiving a prosthetic limb via Medicare, a US-based, federal health insurance program.[37] Additionally, through legal and ethical analyses of the reimbursement process in Germany, it was found that the issue of coverage requires "further empirical investigation, a stakeholder dialogue, and maybe even policy changes."[38]

A myoelectric prosthesis may cost between $20,000 and $30,000,[35] and an AI prosthesis typically costs around $190,000.[33] Whilst this study is from 2016, it still serves to exemplify the large difference in cost for AI-powered devices. Thus, digital prostheses remain far more expensive, often making them a luxury item rather than an essential medical device due to their advanced functionality and extensive R&D requirements. However, there have been recent changes in Clinical Commissioning Policies in the UK's National Health Service (NHS) to support multi-grip myoelectric prosthetic hands for eligible patients with congenital or acquired upper limb loss.[39]

All prosthetic types face ongoing re-purchasing and maintenance costs. In addition, due to growth, children need entirely new prosthetic devices on a short cadence.[40] These additional costs following the initial investment can accumulate, causing the price of prosthetic maintenance and replacement to become a lifelong burden rather than a one-off purchase.

Emotional Benefits and Social Preference:

Limb loss can be a traumatic experience;38 therefore, the purpose of a prosthesis is often twofold: functionally replace the missing limb and emotionally support the patient in recovering from their limb loss.[41] Social reactions from the user's environment strongly influence user satisfaction. Both the emotional benefits and social preference significantly affect how a prosthesis affects a user's self-image, social interactions, and quality of life (QOL).[38]

One study found that prosthesis use can provide ease of social interaction and evade forms of infantilization, indicating that regular use can lead to an amputee feeling self-sufficient.[42] Another study reported numerous positive lifestyle changes associated with prosthesis use, further evidencing its emotional benefits.[43] A study in Saudi Arabia found a positive correlation between psychological adjustment to the prosthesis and the physical component summary and mental component summary of QOL.[44] Similarly, an international multicenter study of QOL and reasons for abandonment in upper limb prostheses produced similar results; the mean QOL (measured using the EQ-5D) was significantly higher in prosthesis users than non-users.[45]

While prosthetic limbs help some 'blend in' with their social environment, many prefer to cover or conceal their prosthesis for social conformity, as found in a study conducted by Murray.[42,46] Visible bionic devices can evoke negative reactions among others, such as a lack of understanding or pity. However, positive reactions are also common, such as 'technical fascination' or interest. Whilst positive reactions exist, many digital prosthesis users describe a negative "dual identity" when wearing their device.[43] Traditional passive prostheses aim to appear natural and human-like; however, traditional body-powered prostheses can require more pronounced and, in some cases, unnatural movement.[47] Therefore, they may not be favored by users in public environments. In such cases, digital prostheses can have increased social preference. Advancements in control technology and algorithms will only increase the 'natural' feel of these devices.[48]

Overall, while digital devices can provide increased emotional benefits through superior function, their "cyborg"-like appearance (Figure 2) can lead to rejection. For traditional prostheses, whilst body-powered prostheses lack functionality and can require unnatural movement for control, passive restorations can most closely resemble human anatomy.

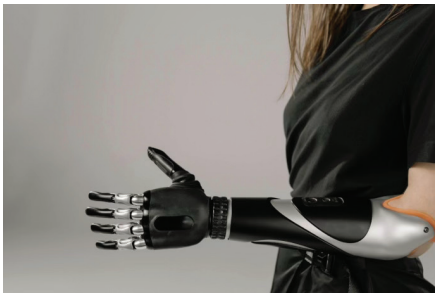


Figure 2: Digital prostheses can have a "cyborg"-like appearance.[49]

DOI: 10.36838/IJHSR819.19

Durability, Robustness, and Simplicity :

Prosthetic durability is the ability to withstand wear and damage during use, while robustness refers to maintaining good condition over its lifetime. Simplicity refers to how easily a new user can understand how to use and operate the device.

Traditional prostheses function fairly reliably in outdoor environments and are less sensitive to water, dirt, and extreme temperatures.[50] They are often more durable and rugged, making them suitable for users who operate in environments where their prosthesis may be susceptible to damage. In contrast, due to their electrical components, digital prostheses, particularly those with microprocessor control, are highly vulnerable to wet environments.[51]

Digital prostheses have the added complication of charging, leading to downtime and inconvenience, which is avoided in traditional prostheses. Therefore, the passive nature of traditional prostheses makes them more suitable in low-resource settings, including post-disaster areas, resource-deprived regions, and areas of war.[52]

Traditional prostheses often excel in simplicity, mostly relying on straightforward, body-powered movement. In contrast, digital prostheses require more complex control skills to be gained through user training. Consequently, digital prosthetic use may not be viable in social groups with low digital literacy rates (such as the elderly, young children, or people in disadvantaged areas). Even for the digitally literate, the learning curve associated with developing the skills required to manipulate digital prostheses can still require extensive training.[53]

Overall, traditional prostheses are less susceptible to damage. Digital prostheses have the impediment of charging, and their complexity can limit accessibility for the less technologically literate. However, digital prostheses have several dedicated training programs available to mitigate this.[54,55]

Discussion

This study attempted to critically compare the three prosthetic types across five user-focused and function-related factors. The main goal is to establish the key differences between each of the prosthetic types to provide a reference for clinicians and users to make more informed choices when selecting a prosthesis. In addition, key future research directions are covered, with the goal of aligning expectations and areas for improvement. Each section is summarized in a traffic light table to make it easy to see the comparisons highlighted in each section and which prosthetic type is the best in each category.

Control and Comfort:

AI-powered prosthetic limbs are typically preferred in terms of comfort due to their harness-free control, variable grip strength, and automatic mode switching. In contrast, traditional prostheses are either body-powered or passive; for the former, it is difficult to avoid the use of harnesses. However, advances in custom-fit harnessing and new textile materials are improving comfort and hygiene in body-powered models.[56]

As mentioned previously, a key advantage of AI-powered prostheses over myoelectric ones is the fact that ML classification systems eliminate manual mode switching for AI-powered prostheses, which is far more convenient than their myoelectric counterparts. Challenges exist in developing AI algorithms for practical use in prostheses. Many factors vary between patients, such as mobility, muscle contraction force, variations in limb position, etc. Therefore, solutions that tailor themselves to patients are required to ensure usability between different patients.[57] Delay remains a major inconvenience for all digital prostheses, as increased latency can lead to a sluggish user experience and higher abandonment rates. From this perspective, one study used high-density electromyography and a convolutional neural network to adapt itself to unique user voluntary muscle contraction patterns,[48] leading to a reduced latency of <116 ms. Such optimization techniques are crucial to ensure widespread adoption of digital prostheses.

Further evaluation is required of materials and components in myoelectric and AI-powered digital prostheses, as their high weight can burden users. Progress includes lightweight actuators for application in both upper- and lower-body prostheses.[58,59] The Össur i-Limb Quantum exemplifies material innovation for lighter prostheses, using titanium to achieve a weight of only 658 grams.[60]

Furthermore, digital prostheses can adapt to muscle signals, allowing users to perform a wider range of tasks compared to traditional prostheses, with a more natural and intuitive feel. However, extensive training is required for patients to manipulate muscles in the residual limb with such precise control. Lack of control is a key factor in prosthetic abandonment, as evidenced by Biddiss and Chau's study that summarized 25 years of publications, finding that the rejection rate of adult prostheses was 26% for body-powered and 23% for myoelectric, indicating that myoelectric control can slightly increase adherence.[61,62] However, a much smaller, more recent study reported that non-use rates can be higher for myoelectric users compared to those using traditional prostheses, which is worth noting. It was found that 44% of 25 users rejected their device, with most complaining about comfort and weight.[63] This could be a consequence of the aforementioned delay caused by EMG signal processing;[64] research suggests an optimal delay to be approximately 100–125 ms.[65] Table 1 provides a summary of the trade-offs between different prosthesis types in terms of control and comfort. Over the five criteria, AI-powered prostheses are superior, followed by myoelectric, and finally traditional.

Table 1: Summary of control and comfort: traditional body-powered systems may require potentially uncomfortable harnesses and offer limited grip-strength variation and modes. Myoelectric and AI-powered systems eliminate harnesses and improve control capabilities, with AI enabling fine-grained grip modulation and automatic mode switching, though both can introduce delays from EMG processing and may increase device mass.

Control and Comfort			
	Traditional	Myoelectric	AI-Powered
Harness	can include uncomfortable harnesses	no harness	no harness
Mass	light	moderate	heavy
Grip Strength	no/limited variation in grip strength	variation in grip strength with training	fine-grained variation in grip strength without training required
Delay	no delay	delays due to EMG signal processing	delays due to EMG signal processing
Modes	no modes available/require terminal devices	need to manually switch modes	automatic mode switching possible

Adaptability :

Overall, AI-powered prosthetic limbs have superior adaptability due to their intuitive nature and support through training programs. Furthermore, machine-learning algorithms enable automatic mode switching and continuous adaptation to users' muscle signals. In contrast, traditional prostheses are simplistic in terms of versatility and control in daily tasks. Currently, terminal devices are employed for different activities; however, switching them out can be tedious. Alternative adaptability methods for traditional prostheses include a quick-disconnect wrist unit to allow for faster switching of terminal devices.[66]

Furthermore, the development of improved training programs could help traditional prosthetic users adapt to their device more effectively. For example, mobile-game-based prosthetic training methods have already been developed for myoelectric prostheses,[38] improving control and user motivation. Translating similar game-based or interactive methods to traditional prostheses could be valuable, as supported by a study on the effects of targeted training for body-powered prostheses, which showed improved functional performance in all users.[67] Another example includes a study conducted on a group of thirty-two able-bodied participants using body-powered bypass prostheses with an augmented reality (AR) training game, which helped participants improve their control and speed.[68]

Alongside training improvements, advancements in traditional prosthetic design are improving adaptability through modular components. An anatomic study of sixty-five cadaveric humeri enabled the creation of a modular prosthesis system, which allowed users to mix and match components, creating prostheses that were customizable despite not being completely bespoke.[69] This modular system could extend to other prosthetic types, facilitating cheaper production by reducing detailed personalization while still maintaining an element of user-specific tailoring.

Table 2: Summary of adaptability across prosthesis types. As the devices can adapt to users, improving control over time, AI-powered devices are superior in adaptability. This is furthered by automatic mode switching, allowing prostheses to adjust their behavior to user actions automatically, rather than using manual switching or terminal devices. Reliance on battery life can be insignificant due to single charges being sufficient for a day's use.

Adaptability			
	Traditional	Myoelectric	AI-Powered
Modes	require terminal devices	uses manual mode switching	uses automatic mode switching
Adapting to the user over time	no adaptation over time	calibrated to muscle signals to perform more varied tasks	adapt and learn from muscle signals over time to perform more varied tasks
Reliance on Battery Life	no batteries required	highly reliant on battery life	highly reliant on battery life

Digital prostheses heavily depend on sensors, power, and battery life, which can pose major safety risks if power fails. Therefore, some advancement in the battery longevity of prostheses should be explored to minimize this issue. The dependence on manufacturers also presents challenges. For example, if a company goes out of business or stops updating a prosthetic model, it can lead to a digital limb becoming useless.[70] Planned obsolescence has become increasingly common in healthcare;[71] this can be an ethical concern, especially for users who are not financially capable of affording updates or new parts. People with disabilities are often already vulnerable to poverty; therefore, re-purchasing additional parts over time is often not feasible.[33] Table 2 provides a summary of the trade-offs between different prosthesis types in terms of adaptability. Over the three criteria, AI-powered prostheses are superior, followed by myoelectric, and finally traditional.

Cost and Accessibility:

Overall, traditional prostheses are the most cost-effective and accessible, especially in resource-deprived regions. However, due to persistent insurance reluctance in prosthetic coverage,[72] even traditional prostheses can be financially inaccessible for many, making digital alternatives even less feasible. Thus, the replacement of traditional prostheses with digital alternatives is largely inhibited by the prohibitive pricing of the industry. The World Health Organization's Standards for Prosthetics and Orthotics has recommended a tiered pricing model to solve this, tailored to varied needs and country resources.[73]

New manufacturing methods, such as 3D printing, are being tested to increase accessibility by significantly reducing costs. This approach produces designs that can better develop affinities for users through an iterative design approach.[74] By 3D scanning the residual limb and importing this data into CAD software, we can mitigate the need for excessive labor and allow prosthetists to focus their attention on where it is truly needed — customization. Additive manufacturing allows rapid prototyping, allowing for edits and customization to be carried out more quickly, even possibly at the point of care within the hospital.[75] The combined effects of reduced labor and improved speed of manufacture can reduce costs. These methods of manufacture have been applied to myoelectric contexts, where the "Mark 5 ™, Mandro" 3D-printed myoelectric prosthetic hand demonstrated that an affordable, reliable, and durable design can have a "major social impact on patients."[76]

The simplicity of traditional prostheses makes them more suitable for users with lower competence due to age or other ability-related factors. Therefore, traditional prostheses often do not have structured support, and further research should be conducted into developing structured training programs to help users adapt to their traditional prostheses, particularly in terms of proprioception and agility. Some developments have been made in this area, with studies showing that designated training programs for traditional prosthetic users can help improve control and speed.[67] Digital prostheses require more training, but the existence of structured training programs helps users adapt more effectively to their devices. Furthermore, since AI-powered prostheses learn the user's behavior over time,[29] they could have a less steep learning curve compared to myoelectric prostheses. Game-based and AR training methods have been proposed to make myoelectric skill development more engaging and accessible outside clinical settings, helping users become accustomed to their less intuitive nature.[51,77]

For those who are technologically illiterate or challenged, such as the elderly, there is a question of whether developing the digital literacy required to operate a digital prosthesis, as

DOI: 10.36838/IJHSR819.19

well as enduring the learning curve, is worth the effort. In addition, low-income countries with lower rates of education often have lower rates of digital literacy.[78] This creates a clear barrier for digital prostheses; more rugged, simple, traditional prostheses may be more accessible to the greater population. Table 3 provides a summary of the trade-offs between different prosthesis types in terms of cost and accessibility. Over the three criteria, traditional prostheses are superior, followed by myoelectric, and finally AI-powered.

Table 3: Summary of cost and accessibility across prosthesis types. Traditional prostheses are the cheapest option and most commonly adopted. The moderate learning curve, while not supported by the learning programs that myoelectric and AI models have, can be easily facilitated by regular use.

Cost and Accessibility			
	Traditional	Myoelectric	AI-Powered
Cost	cheapest	more expensive	most expensive
Accessibility	most accessible	less accessible	least accessible
Learning Curve	moderate learning curve, without support	steep learning curve, with support	moderate learning curve, with support

Emotional Benefits and Social Preference:

Overall, traditional prostheses offer greater emotional benefits, user social preference, and social conformity; the bionic nature of digital options can be discomforting, especially in rural contexts where acceptance of amputation and prosthetic usage is low.

Users of all types of prostheses experience some level of improvement in QOL; however, the "cyborg-like "dual identity" created by digital prostheses[43] can be harmful to a person's mental state and self-perception. It is, therefore, important to consider both the positive and unintended negative social consequences of the augmentation of technology in digital prostheses[79] and perhaps biomedical devices as a whole. This could be changed by making digital prostheses more human-like, such as adopting human skin tones and texture for the unit. For example, Ossur's i-Limb skin match is a silicone glove that fits over the i-Limb model, which has a very realistic appearance, concealing the myoelectric prosthesis while maintaining function.[80]

For many users, the ability to conceal their prosthesis is critical for their social conformity.[42] This is fairly difficult for digital prostheses, where their bionic nature is highly visible. However, for traditional prostheses, their appearance is much simpler to conceal, especially for passive prostheses, which are intended to be non-functional but cosmetically identical to a real limb.

The interest or fascination that some outsiders feel when interacting with a digital prosthetic user can have a positive effect on the user themselves.[43] While this interest may be prevalent in Western cultures, the digitization of prosthetics may not be favored in rural contexts, such as in Africa or South Asia, due to more traditional beliefs. This is exemplified by a systematic review, which noted that cultural barriers, including the traditional beliefs of Sub-Saharan Africa, hindered the use of prostheses.[81] To ameliorate this, awareness efforts are crucial for rural populations to educate them about the capabilities and benefits of digital prostheses and prosthetics as a whole. This could have some impact in increasing acceptance; however, aiming to change systemic traditional values and opinions can be unfeasible.

Incorporating patient perspectives into prosthetic design can help address social sources of dissatisfaction. Individual choice is both emotionally beneficial to users and is necessary due to the individuality of user needs and wants for prosthetics. Integrating user input into prosthetic design enables the incorporation of social factors into Clinical Practice Guidelines, helping reduce the negative social impacts of prosthetic use.[38] Table 4 provides a summary of the trade-offs between different prosthesis types in terms of emotional benefits and social preference. Over the four criteria, traditional prostheses are superior, followed by myoelectric and AI-powered prostheses.

Table 4: Summary of emotional benefits and social preference. AI-powered and myoelectric devices' advanced function lends a higher level of independence and creates interest among others. However, this fascination may be limited to Western contexts, where advancement in human body-related technology is accepted to a greater degree than in others. Additionally, traditional prostheses can usually be easily concealed to evade stigma.

Emotional Benefits and Social Preference			
	Traditional	Myoelectric	AI-Powered
Improvement to QOL	general improvement	general improvement, but increased 'disability-related feelings of stigma'	general improvement, but increased 'disability-related feelings of stigma'
Level of independence provided	moderate to low	high	high
Ability to conceal prosthesis	high for most traditional prostheses	low	low
Technical fascination	low	high, but may be limited to Western contexts	very high, but may be limited to Western contexts

Durability, Robustness, and Simplicity:

Myoelectric and AI-powered prostheses have low resistance to adverse conditions, particularly water, whereas traditional prostheses are less sensitive to weather and terrain. Research into water-resistant digital prostheses is ongoing, with a study developing an EMG sensor that can be used underwater.82 It maintained EMG signal quality above and completely submerged below water, evidencing the potential for less sensitive digital prostheses in the future.

Furthermore, charging requirements can inconvenience digital prosthetic users and limit viability in low-resource settings with scarce electricity. Efforts to reduce the energy requirements of prostheses include a study in the context of a powered prosthetic ankle; using a convex optimization formulation, they designed the stiffness of series elastic actuators to minimize energy consumption.[83] Table 5 provides a summary of the trade-offs between different prosthesis types in terms of durability, robustness, and simplicity. Over the three criteria, traditional prostheses are superior, followed by myoelectric and AI-powered.

Table 5: Summary of durability, robustness, and simplicity. Traditional prostheses' low complexity and therefore resistance to adverse conditions allow use in many climates and environments. Their lack of need to charge maintains simplicity.

Durability, Robustness and Simplicity			
	Traditional	Myoelectric	AI-Powered
Complexity	least complex	more complex	more complex
Resistance to Adverse Conditions	high resistance	low resistance	low resistance
Charging	does not require charging	requires charging	requires charging

Integrative Perspective:

While the five factors have been discussed separately in this study, it is important to acknowledge that complex relationships exist among them. Therefore, this section provides an integrative perspective on these relationships, with the aim of

allowing a user to identify which prosthesis is best suited to their needs.

Training burden affects control, comfort, accessibility and can cause prosthetic abandonment. Issues can arise when users transfer between clinics, and training programs are not consistent. Therefore, training programs should be mandatory (perhaps by government legislation, with potential reimbursement for time and effort) and device-specific (designed by the prosthetic company itself, thus preventing disruption of continuity of care).

Furthermore, user preference cannot be determined by prosthetic function alone, as all factors interact uniquely for each patient. User selection should not be limited by clinic-specific collections of prosthetic devices; users should be able to self-evaluate their preferences and assess all available options before selection. There is currently no standardized way to compare all types of prosthetic modes (traditional, myoelectric, and AI-powered) across economic, functional, and social differences. Studies such as this one are examples of current solutions to this issue; however, such papers are not approachable for public consumption. Therefore, a centralized, patient-facing, online prosthesis comparison platform (with integrated user opinions, product specifications, costs, etc.) can be an accessible way for users to decide which model would be best for their unique needs.

In addition, AI-powered devices have the hidden cost of pre-training through calibration systems, as a brand-new device will not be trained to the specific user. Since the ML algorithms must adapt to each user, this comes at an additional cost.

Research trends are directed towards advancing functionality and control in these devices; however, this innovation can often increase the cost of devices. Innovation and affordability are viewed as separate goals in the industry. Future research should consider whether advanced function can be developed alongside cost reduction strategies to maximize accessibility.

Accessibility is not only financial but also technological. Even where funding is available, devices that require charging, updates, calibration, or high digital literacy may be unsuitable for some users. A solution could involve adapting training programs or offering alternate training programs to accommodate those with lower digital literacy, allowing them to access these more advanced digital devices.

Overall, from the user's perspective, considerations for both upper and lower-limb devices were fairly generalizable to both – have an agreeable appearance, be accepted socially, and function well enough to maintain QOL. The considerations are also mostly generalizable from a clinician's perspective, but the design and manufacturing considerations are different, particularly since lower-limb devices are load-bearing. Prosthetists would use different methods to make devices comfortable and functional. Table 6 summarizes all five factors to provide a succinct summary of this review's findings and can be used as a ready reference for selection.

Current research directions revolve around reducing costs through point-of-care and additive manufacturing; reducing EMG signal processing delays; considering alternative materials to improve device weight and cost; standardizing training programs, ensuring they are adaptable to users of all digital literacy levels; and improving social acceptance through user-centered prosthetic design.[84-87]

Table 6: Summary of all five factors. Depending on client choice, socioeconomic conditions, and clinical settings, appropriate prosthesis types can be selected. AI-powered prostheses provide the best control and comfort, and the highest adaptability, but are the heaviest, most complex, battery-dependent, and least accessible due to cost. Traditional prostheses are most affordable, robust, simple, and widely accessible but offer the least control, adaptability, and independence. Myoelectric prostheses stand as an intermediate by improving function over traditional devices but retaining training burden, battery reliance, and reduced robustness, despite remaining more accessible than AI-powered systems.

Cornerstone			
	Traditional	Myoelectric	AI-Powered
Control and Comfort	Simple control	Advanced control, unintuitive but supported (through training programs)	Advanced control, intuitive and supported (through training programs and ML algorithms)
	Least comfortable	More comfortable	Most comfortable
Adaptability	No modes, binary states	Manual mode switching	Automatic mode switching
	Doesn't adapt to user over time	Doesn't adapt to user over time	Adapts and learns from muscle signals over time to perform more varied tasks
	No reliance on battery life	Reliant on battery life	Reliant on battery life
Cost and Accessibility	Cheapest	Expensive	Most expensive
	Most accessible	Less accessible	Least accessible
	Moderate learning curve, without support	Steep learning curve, with support	Moderate learning curve, with support
Emotional Benefits and Social Preference	General improvement to QOL	General improvement to QOL, but increased 'disability-related feelings of stigma'	General improvement to QOL, but increased 'disability-related feelings of stigma'
	Moderate to low independence	High independence	High independence
	Some types are easy to conceal in public, others are not	Difficult to conceal in public	Difficult to conceal in public
	No technical fascination	High technical fascination, may be limited to Western contexts	High technical fascination, may be limited to Western contexts
Durability, Robustness and Simplicity	Least complex	More complex	More complex
	High resistance to adverse conditions	Low resistance to adverse conditions	Low resistance to adverse conditions
	Doesn't require charging	Requires charging	Requires charging

Case Study:

To exemplify the comparisons made in this paper, three flagship devices, one of each prosthetic type, have been identified and compared across the five factors.

Table 7: The Hosmer 5X Hook, Ottobock Michelangelo Hand, and the Coapt Complete Control Gen2 systems have been compared across the five factors highlighted in the paper.

Type	Traditional	Myoelectric	AI-Powered
Product Name	Hosmer 5X Hook	Ottobock Michelangelo Hand	Coapt Complete Control Gen2 + A compatible prosthetic hand[88]
Control and Comfort	Body-powered cable/harness control.	Two electrodes capture contraction of forearm muscles.[89]	Uses two dome electrodes with 8 EMG channels. Continuously adapts to user through 'Adaptive Advance ®' and 'Control Coach ®' calibration systems.
Adaptability	Binary state – open or closed.[90]	7 different grips possible through particular finger alignment and motorized thumb.[89]	Uses a continuous-learning algorithm, 'Adaptive Advance ®' that combines previous and new calibration data to adapt and learn to the user's natural changes in control over time.
Cost and Accessibility	$679.90 (terminal device) only.[91]	~$60,000-75,000 (prices not available publicly). Ottobock has developed accompanying exercises to train user familiarity (mandatory completion prior to use).[92]	~$70,000-100,000+ (prices not available publicly). Uses a proprietary AI model 'Control Coach ®' to analyze user's calibration data to determine which types of movement will get the least reliable control. Then suggests specific calibration exercises which will add to existing data to allow for better control. This supports user training.
Emotional Benefits and Social Preference	Fairly concealable. Hook however does not mimic human anatomy at all.	Anatomically mimics hand. Can purchase cosmetic, skin-color silicone cover to conceal.[93]	Dependent on which compatible hand you select to use with the control system.
Durability, Robustness, and Simplicity	Waterproof up to 1 meter.[94]	Must be protected from dirt and humidity. Requires charging.[92]	Sensitive to 'extreme temperatures' (<-20°C, >45°C) and excessive moisture and dust. Requires charging.

DOI: 10.36838/IJHSR819.19

Limitations:

This study is not without limitations. Firstly, the studies referenced here are slightly biased towards upper-limb prostheses, so their findings may not fully apply to lower-limb or mixed amputation contexts. This is because myoelectric/AI-powered systems are more prevalent in upper-limb devices. Secondly, many of the studies are conducted in laboratory settings, and the findings may not fully translate to the real world. For example, cost reductions achieved through specific fabrication methods could be offset by higher insurance mark-ups and changing governance, increasing the final selling price to users. Additionally, prosthesis choice is inevitably controlled by the individual needs of each user. However, we did not suggest which type of prosthesis may be preferable for different user groups (e.g., age, time since amputation, amputation type, user needs). The information in this paper is only presented as a screening tool and is not intended to replace clinician decision-making entirely. Factors such as intended function, age, demographic, etc. should all be taken into account by the treating clinician. Finally, some of the studies referenced here are not recent (median value = 2020), and many studies are conducted in Western contexts. This leads to a bias towards Western populations, which is particularly relevant to *Sec. Emotional Benefits and Social Preference*, where attitudes and beliefs towards prosthetic use differ significantly between Western contexts and the Global South. Furthermore, more recent technologies may be underrepresented.

■ Conclusion

This paper holistically compared traditional, myoelectric, and AI-powered prostheses on five different factors, spanning function, use, economic impact, social impact, and longevity.

The findings of this paper showed the following:

1. Traditional prostheses: These are the cheapest option and are currently most widely accepted, particularly in rural environments. They are most easily adoptable, due to being the least technically complex and having the strongest durability against adverse conditions. However, they provide the least independence.

2. Myoelectric prostheses: Myoelectric prostheses form an intermediate ground between the simpler traditional prostheses and higher-complexity AI-powered ones. They are better than the former in terms of control but are also more expensive and less accessible; in comparison to the latter, they have less advanced control but are more accessible and cheaper.

3. AI-powered prostheses: In comparison to the other prosthesis types, they excelled in control and comfort due to their advanced control algorithms and had superior adaptability due to their ability to adapt to users. They had some strength in emotional benefits, mainly due to their advanced function, allowing greater QOL. These devices have the greatest potential for development, as current research is primarily directed towards their improvement and advancement.

Patients may lean towards traditional prostheses when access and costs are a concern, but they prefer myoelectric and AI-powered ones when control and independence are of greater concern. Future research should focus on cost-saving efforts (e.g., greater use of additive manufacturing, on-site manufacturing, alternate material usage, etc.) to increase access; improved insurance and governance policies; optimized control through AI-driven algorithms; faster EMG signal processing; and weight reduction through material choice.

■ Acknowledgments

Thank you to Vishal Pendse for his guidance and mentorship during the development of this literature review. His support was invaluable to the research process.

■ References

1. Sansoni, S.; Speer, L.; Wodehouse, A.; Buis, A. Aesthetic of Prosthetic Devices: From Medical Equipment to a Work of Design. In *Emotional Engineering Volume 4*; Fukuda, S., Ed.; Springer International Publishing: Cham, 2016; pp 73–92. https://doi.org/10.1007/978-3-319-29433-9_5.
2. Engdahl, S. M.; Gonzalez, M. A.; Lee, C.; Gates, D. H. Perspectives on the Comparative Benefits of Body-Powered and Myoelectric Upper Limb Prostheses. *J. NeuroEngineering Rehabil.* **2024**, *21* (1), 138. https://doi.org/10.1186/s12984-024-01436-4.
3. O'Keeffe, B. Prosthetic Rehabilitation of the Upper Limb Amputee. *Indian J. Plast. Surg.* **2011**, *44* (2), 246. https://doi.org/10.4103/0970-0358.85346.
4. Maat, B.; Smit, G.; Plettenburg, D.; Breedveld, P. Passive Prosthetic Hands and Tools: A Literature Review. *Prosthet. Orthot. Int.* **2018**, *42* (1), 66–74. https://doi.org/10.1177/0309364617691622.
5. Hafner, B. J.; Willingham, L. L.; Buell, N. C.; Allyn, K. J.; Smith, D. G. Evaluation of Function, Performance, and Preference as Transfemoral Amputees Transition From Mechanical to Microprocessor Control of the Prosthetic Knee. *Arch. Phys. Med. Rehabil.* **2007**, *88* (2), 207–217. https://doi.org/10.1016/j.apmr.2006.10.030.
6. Sensinger, J. W.; Hill, W.; Sybring, M. Prostheses—Assistive Technology—Upper. In *Encyclopedia of Biomedical Engineering*; Elsevier, **2019**; pp 632–644. https://doi.org/10.1016/B978-0-12-801238-3.99912-4.
7. Elbasiouny, S. M. The Neurophysiology of Sensorimotor Prosthetic Control. *BMC Biomed. Eng.* **2024**, *6* (1), 9. https://doi.org/10.1186/s42490-024-00084-y.
8. Kaufman, K. R.; Levine, J. A.; Brey, R. H.; McCrady, S. K.; Padgett, D. J.; Joyner, M. J. Energy Expenditure and Activity of Transfemoral Amputees Using Mechanical and Microprocessor-Controlled Prosthetic Knees. *Arch. Phys. Med. Rehabil.* **2008**, *89* (7), 1380–1385. https://doi.org/10.1016/j.apmr.2007.11.053.
9. Hahne, J. M.; Wilke, M. A.; Koppe, M.; Farina, D.; Schilling, A. F. Longitudinal Case Study of Regression-Based Hand Prosthesis Control in Daily Life. *Front. Neurosci.* **2020**, *14*, 600. https://doi.org/10.3389/fnins.2020.00600.
10. Heerschop, A.; Van Der Sluis, C. K.; Otten, E.; Bongers, R. M. Looking beyond Proportional Control: The Relevance of Mode Switching in Learning to Operate Multi-Articulating Myoelectric Upper-Limb Prostheses. *Biomed. Signal Process. Control* **2020**, *55*, 101647. https://doi.org/10.1016/j.bspc.2019.101647.
11. Clancy, E. A.; Martinez-Luna, C.; Wartenberg, M.; Dai, C.; Farrell, T. R. Two Degrees of Freedom Quasi-Static EMG-Force at the Wrist Using a Minimum Number of Electrodes. *J. Electromyogr. Kinesiol.* **2017**, *34*, 24–36. https://doi.org/10.1016/j.jelekin.2017.03.004.
12. Luu, D. K.; Nguyen, A. T.; Jiang, M.; Drealan, M. W.; Xu, J.; Wu, T.; Tam, W.; Zhao, W.; Lim, B. Z. H.; Overstreet, C. K.; Zhao, Q.;

Cheng, J.; Keefer, E. W.; Yang, Z. Artificial Intelligence Enables Real-Time and Intuitive Control of Prostheses via Nerve Interface. *IEEE Trans. Biomed. Eng*. **2022**, 69 (10), 3051–3063. https://doi.org/10.1109/TBME.2022.3160618.

13. Joshi, D. C.; Kumar, P.; Joshi, R. C.; Mitra, S. AI-Enhanced Analysis to Investigate the Feasibility of EMG Signals for Prosthetic Hand Force Control Incorporating Anthropometric Measures. *Prosthesis* **2024**, *6* (6), 1459–1478. https://doi.org/10.3390/prosthesis6060106.
14. Chopra, S.; Emran, T. B. Advances in AI-Based Prosthetics Development: Editorial. *Int. J. Surg.* **2024**, *110* (8), 4538–4542. https://doi.org/10.1097/JS9.0000000000001573.
15. Engdahl, S. M.; Gates, D. H. Differences in Quality of Movements Made with Body-Powered and Myoelectric Prostheses during Activities of Daily Living. *Clin. Biomech*. **2021**, *84*, 105311. https://doi.org/10.1016/j.clinbiomech.2021.105311.
16. Walters, S.; Seminati, E.; Metcalfe, B.; Bailey, N. Y.; Pegg, E. C. Demystifying Upper Limb Hybrid Prostheses-a Scoping Review. *Front. Rehabil. Sci*. **2025**, *6*, 1610336. https://doi.org/10.3389/fresc.2025.1610336.
17. Guðfinnsdóttir, Þ. Design and Evaluation of a New Harnessing System for Body-Powered Prostheses. MSc Thesis, TU Delft University of Technology, 2013. https://resolver.tudelft.nl/uuid:91338545-0a2d-47ee-b9bf-22236d5242b8.
18. Dewald, H. A.; Lukyanenko, P.; Lambrecht, J. M.; Anderson, J. R.; Tyler, D. J.; Kirsch, R. F.; Williams, M. R. Stable, Three Degree-of-Freedom Myoelectric Prosthetic Control via Chronic Bipolar Intramuscular Electrodes: A Case Study. *J. NeuroEngineering Rehabil.* **2019**, *16* (1), 147. https://doi.org/10.1186/s12984-019-0607-8.
19. Berning, J.; Francisco, G. E.; Chang, S.-H.; Fregly, B. J.; O'Malley, M. K. Myoelectric Control and Neuromusculoskeletal Modeling: Complementary Technologies for Rehabilitation Robotics. *Curr. Opin. Biomed. Eng.* **2021**, *19*, 100313. https://doi.org/10.1016/j.cobme.2021.100313.
20. Chen, Z.; Min, H.; Wang, D.; Xia, Z.; Sun, F.; Fang, B. A Review of Myoelectric Control for Prosthetic Hand Manipulation. *Biomimetics* **2023**, *8* (3), 328. https://doi.org/10.3390/biomimetics8030328.
21. Vilarino, M.; Moon, J.; Rogner Pool, K.; Varghese, J.; Ryan, T.; V. Thakor, N.; Kaliki, R. Outcomes and Perception of a Conventional and Alternative Myoelectric Control Strategy: A Study of Experienced and New Multiarticulating Hand Users. *JPO J. Prosthet. Orthot*. **2015**, *27* (2), 53–62. https://doi.org/10.1097/JPO.0000000000000055.
22. Burtsev, N. I.; Shagdurov, V. C.; Demkin, I. O. Application of the Random Forest Machine Learning Algorithm for Recognizing Patient Arm Movements While Using a Bionic Prosthesis; St. Petersburg, Russia, 2019; p 020010. https://doi.org/10.1063/1.5121935.
23. Biddiss, E.; Chau, T. Upper-Limb Prosthetics: Critical Factors in Device Abandonment. *Am. J. Phys. Med. Rehabil*. **2007**, *86* (12), 977–987. https://doi.org/10.1097/PHM.0b013e3181587f6c.
24. Andrade, A. O.; Pereira, A. A.; Walter, S.; Almeida, R.; Loureiro, R.; Compagna, D.; Kyberd, P. J. Bridging the Gap between Robotic Technology and Health Care. *Biomed. Signal Process. Control* **2014**, *10*, 65–78. https://doi.org/10.1016/j.bspc.2013.12.009.
25. Lee, K. H.; Bin, H.; Kim, K.; Ahn, S. Y.; Kim, B.-O.; Bok, S.-K. Hand Functions of Myoelectric and 3D-Printed Pressure-Sensored Prosthetics: A Comparative Study. *Ann. Rehabil. Med*. **2017**, *41* (5), 875. https://doi.org/10.5535/arm.2017.41.5.875.
26. Fryer, C.; Michael, J. Upper-Limb Prosthetics: Body-Powered Components. In *Atlas of Limb Prosthetics: Surgical, Prosthetic, and Rehabilitation Principles*; Mosby Year Book: St. Louis, MO, 1992; pp 173–232.
27. Hussaini, A.; Kyberd, P.; Mulindwa, B.; Ssekitoleko, R.; Keeble, W.; Kenney, L.; Howard, D. 3D Printing in LMICs: Functional Design for Upper Limb Prosthetics in Uganda. *Prosthesis* **2023**, *5* (1), 130–147. https://doi.org/10.3390/prosthesis5010011.
28. Widehammar, C.; Hiyoshi, A.; Lidström Holmqvist, K.; Lindner, H.; Hermansson, L. Effect of Multi-Grip Myoelectric Prosthetic Hands on Daily Activities, Pain-Related Disability and Prosthesis Use Compared with Single-Grip Myoelectric Prostheses: A Single-Case Study. *J. Rehabil. Med.* **2021**. https://doi.org/10.2340/jrm.v53.807.
29. Pilarski, P. M.; Dawson, M. R.; Degris, T.; Carey, J. P.; Chan, K. M.; Hebert, J. S.; Sutton, R. S. Adaptive Artificial Limbs: A Real-Time Approach to Prediction and Anticipation. *IEEE Robot. Autom. Mag.* **2013**, *20* (1), 53–64. https://doi.org/10.1109/MRA.2012.2229948.
30. Mitchell, M.; Gorelick, M.; Anderson, D. I.; Atkins, D. J. Prosthetic Education: Are Occupational Therapy Students' Needs Being Met? *Open J. Ther. Rehabil.* **2014**, *02* (01), 5–11. https://doi.org/10.4236/ojtr.2014.21002.
31. Rajah, J. K.; Chernicoff, W.; Hutchison, C. J.; Gonçalves, P.; Kopainsky, B. Enabling Mobility: A Simulation Model of the Health Care System for Major Lower-Limb Amputees to Assess the Impact of Digital Prosthetics Services. *Systems* **2023**, *11* (1), 22. https://doi.org/10.3390/systems11010022.
32. Rabbitt, L.; Curneen, J.; Hobbins, A.; Gillespie, P.; Dennedy, M. C. Assessing the Financial Burden on Patients and Their Households Attending Hospital Clinics: A Pilot Cross-Sectional Study. *BMC Health Serv. Res.* **2025**, *25* (1), 1390. https://doi.org/10.1186/s12913-025-13503-0.
33. Kwek, E.; Choi, M. Is a Prosthetic Arm Customized PRADA? A Critical Perspective on the Social Aspects of Prosthetic Arms. *Disabil. Soc.* **2016**, *31* (8), 1144–1147. https://doi.org/10.1080/09687599.2016.1235308.
34. Ferguson, H. J. M.; Nightingale, P.; Pathak, R.; Jayatunga, A. P. The Influence of Socio-Economic Deprivation on Rates of Major Lower Limb Amputation Secondary to Peripheral Arterial Disease. *Eur. J. Vasc. Endovasc. Surg*. **2010**, *40* (1), 76–80. https://doi.org/10.1016/j.ejvs.2010.03.008.
35. McGimpsey, G.; Bradford, T. *Limb Prosthetics Services and Devices: Critical Unmet Need: Market Analysis*; National Institute of Standards and Technology (NIST), 2017. https://www.nist.gov/system/files/documents/2017/04/28/239_limb_prosthetics_services_devices.pdf (accessed 2025-12-08).
36. McLarney, M.; Pezzin, L. E.; McGinley, E. L.; Prosser, L.; Dillingham, T. R. The Prevalence of Lower Limb Loss in Children and Associated Costs of Prosthetic Devices: A National Study of Commercial Insurance Claims. *Prosthet. Orthot. Int.* **2021**, *45* (2), 115–122. https://doi.org/10.1177/0309364620968645.
37. United States Government Accountability Office. *Rehabilitation Services and Outcomes for Medicare Beneficiaries*; GAO-25-106406; United States Government Accountability Office: Washington, D.C, 2024. https://www.gao.gov/assets/gao-25-106406-highlights.pdf (accessed 2025-12-08).
38. Baumann, M. F.; Frank, D.; Kulla, L.-C.; Stieglitz, T. Obstacles to Prosthetic Care—Legal and Ethical Aspects of Access to Upper and Lower Limb Prosthetics in Germany and the Improvement of Prosthetic Care from a Social Perspective. *Societies* **2020**, *10* (1), 10. https://doi.org/10.3390/soc10010010.
39. NHS England. *Clinical Commissioning Policy: Multi-Grip Prosthetic Hand (All Ages)*; NHS England: London, UK, 2022. https://www.england.nhs.uk/wp-content/uploads/2022/09/2009-cli-

DOI: 10.36838/IJHSR819.19

nicial-comm-policy-multi-grip-prosthetic-hand-all-ages-1.pdf (accessed 2025-12-08).
40. Dickinson, A.; Gates, L.; Metcalf, C.; Spurway, C.; Kheng, S.; Heang, T.; Sam, B.; Harte, C.; Simpson, S.; Worsley, P.; Ostler, C.; Donovan-Hall, M.; Channon, A. Understanding Maintenance, Repair, and Replacement of Prosthetic Limbs Using Routinely-Collected Data: A Retrospective Study over Three Decades in Cambodia. *J. Glob. Health* **2025**, *15*, 04135. https://doi.org/10.7189/jogh.15.04135.
41. Volpini Lana, M. R.; Maia, J. P.; Ribeiro, J. P. Current Insights into Lower Limb Prostheses: Brazilian Users' Perceptions on the (Re) Construction of Body Image and Identity. *Ir. J. Occup. Ther.* **2025**, *53* (1), 4–11. https://doi.org/10.1108/IJOT-06-2024-0029.
42. Murray, C. D. The Social Meanings of Prosthesis Use. *J. Health Psychol.* **2005**, *10* (3), 425–441. https://doi.org/10.1177/1359105305051431.
43. Bretschneider, M.; Meyer, B.; Asbrock, F. The Impact of Bionic Prostheses on Users' Self-Perceptions: A Qualitative Study. *Acta Psychol. (Amst.)* **2023**, *241*, 104085. https://doi.org/10.1016/j.actpsy.2023.104085.
44. Alessa, M.; Alkhalaf, H. A.; Alwabari, S. S.; Alwabari, N. J.; Alkhalaf, H.; Alwayel, Z.; Almoaibed, F. The Psychosocial Impact of Lower Limb Amputation on Patients and Caregivers. *Cureus* **2022**. https://doi.org/10.7759/cureus.31248.
45. Yamamoto, M.; Chung, K. C.; Sterbenz, J.; Shauver, M. J.; Tanaka, H.; Nakamura, T.; Oba, J.; Chin, T.; Hirata, H. Cross-Sectional International Multicenter Study on Quality of Life and Reasons for Abandonment of Upper Limb Prostheses. *Plast. Reconstr. Surg. - Glob. Open* **2019**, *7* (5), e2205. https://doi.org/10.1097/GOX.0000000000002205.
46. Widehammar, C.; Pettersson, I.; Janeslätt, G.; Hermansson, L. The Influence of Environment: Experiences of Users of Myoelectric Arm Prosthesis—a Qualitative Study. *Prosthet. Orthot. Int.* **2018**, *42* (1), 28–36. https://doi.org/10.1177/0309364617704801.
47. Engdahl, S. M.; Lee, C.; Gates, D. H. A Comparison of Compensatory Movements between Body-Powered and Myoelectric Prosthesis Users during Activities of Daily Living. *Clin. Biomech.* **2022**, *97*, 105713. https://doi.org/10.1016/j.clinbiomech.2022.105713.
48. Tam, S.; Boukadoum, M.; Campeau-Lecours, A.; Gosselin, B. Intuitive Real-Time Control Strategy for High-Density Myoelectric Hand Prosthesis Using Deep and Transfer Learning. *Sci. Rep.* **2021**, *11* (1), 11275. https://doi.org/10.1038/s41598-021-90688-4.
49. cottonbro studio. *Close-up of a Woman Wearing a Sleek, Modern Prosthetic Arm. Studio Shot Emphasizes Technology and Innovation.*; 2020. https://www.pexels.com/photo/close-up-shot-of-a-woman-with-prosthetic-arm-6153078/ (accessed 2026-06-06).
50. Trejo-Letechipia, M. A.; Rodriguez-Sanchez, D. A.; González-González, R. B.; Perez-Sanpablo, A. I.; Arizmendi-Morquecho, A. M.; Lara-Ceniceros, T. E.; Bonilla-Cruz, J.; Cortes-Ramirez, J. A. Design and Manufacturing of a Body-Powered Hook with Force Regulation System and Composite-Based Nanomaterials. *Appl. Sci.* **2021**, *11* (9), 4225. https://doi.org/10.3390/app11094225.
51. McDonald, C. L.; Kahn, A.; Hafner, B. J.; Morgan, S. J. Prevalence of Secondary Prosthesis Use in Lower Limb Prosthesis Users. *Disabil. Rehabil.* **2024**, *46* (5), 1016–1022. https://doi.org/10.1080/09638288.2023.2182919.
52. McFarland, L. V.; Winkler, S. L. H.; Heinemann, A. W.; Jones, M.; Esquenazi, A. Unilateral Upper-Limb Loss: Satisfaction and Prosthetic-Device Use in Veterans and Servicemembers from Vietnam and OIF/OEF Conflicts. *J. Rehabil. Res. Dev.* **2010**, *47* (4), 299. https://doi.org/10.1682/JRRD.2009.03.0027.
53. Damercheli, S.; Morrenhof, K.; Ahmed, K.; Ortiz-Catalan, M. Performance in Myoelectric Pattern Recognition Improves with Transcranial Direct Current Stimulation. *Sci. Rep.* **2024**, *14* (1), 11744. https://doi.org/10.1038/s41598-024-62185-x.
54. Widehammar, C.; Lidström Holmqvist, K.; Hermansson, L. Training for Users of Myoelectric Multigrip Hand Prostheses: A Scoping Review. *Prosthet. Orthot. Int.* **2021**, *45* (5), 393–400. https://doi.org/10.1097/PXR.0000000000000037.
55. Winslow, B. D.; Ruble, M.; Huber, Z. Mobile, Game-Based Training for Myoelectric Prosthesis Control. *Front. Bioeng. Biotechnol.* **2018**, *6*, 94. https://doi.org/10.3389/fbioe.2018.00094.
56. Vujaklija, I.; Farina, D.; Aszmann, O. New Developments in Prosthetic Arm Systems. *Orthop. Res. Rev.* **2016**, *Volume 8*, 31–39. https://doi.org/10.2147/ORR.S71468.
57. Parajuli, N.; Sreenivasan, N.; Bifulco, P.; Cesarelli, M.; Savino, S.; Niola, V.; Esposito, D.; Hamilton, T. J.; Naik, G. R.; Gunawardana, U.; Gargiulo, G. D. Real-Time EMG Based Pattern Recognition Control for Hand Prostheses: A Review on Existing Methods, Challenges and Future Implementation. *Sensors* **2019**, *19* (20), 4596. https://doi.org/10.3390/s19204596.
58. Yang, H.; Tao, Z.; Yang, J.; Ma, W.; Zhang, H.; Xu, M.; Wu, M.; Sun, S.; Jin, H.; Li, W.; Wang, L.; Zhang, S. A Lightweight Prosthetic Hand with 19-DOF Dexterity and Human-Level Functions. *Nat. Commun.* **2025**, *16* (1), 955. https://doi.org/10.1038/s41467-025-56352-5.
59. Tran, M.; Gabert, L.; Hood, S.; Lenzi, T. A Lightweight Robotic Leg Prosthesis Replicating the Biomechanics of the Knee, Ankle, and Toe Joint. *Sci. Robot.* **2022**, *7* (72), eabo3996. https://doi.org/10.1126/scirobotics.abo3996.
60. Siegel, J. R.; Harwood, J. K.; Lau, A. C.; Brenneis, D. J. A.; Dawson, M. R.; Pilarski, P. M.; Schofield, J. S. A Performance Evaluation of Commercially Available and 3D-Printable Prosthetic Hands: A Comparison Using the Anthropomorphic Hand Assessment Protocol. *BMC Biomed. Eng.* **2024**, *6* (1), 11. https://doi.org/10.1186/s42490-024-00086-w.
61. Yamamoto, M.; Chung, K. C.; Sterbenz, J.; Shauver, M. J.; Tanaka, H.; Nakamura, T.; Oba, J.; Chin, T.; Hirata, H. Cross-Sectional International Multicenter Study on Quality of Life and Reasons for Abandonment of Upper Limb Prostheses. *Plast. Reconstr. Surg. - Glob. Open* **2019**, *7* (5), e2205. https://doi.org/10.1097/GOX.0000000000002205.
62. Biddiss, E. A.; Chau, T. T. Upper Limb Prosthesis Use and Abandonment: A Survey of the Last 25 Years. *Prosthet. Orthot. Int.* **2007**, *31* (3), 236–257. https://doi.org/10.1080/03093640600994581.
63. Salminger, S.; Stino, H.; Pichler, L. H.; Gstoettner, C.; Sturma, A.; Mayer, J. A.; Szivak, M.; Aszmann, O. C. Current Rates of Prosthetic Usage in Upper-Limb Amputees – Have Innovations Had an Impact on Device Acceptance? *Disabil. Rehabil.* **2022**, *44* (14), 3708–3713. https://doi.org/10.1080/09638288.2020.1866684.
64. Wang, X.; Ao, D.; Li, L. Robust Myoelectric Pattern Recognition Methods for Reducing Users' Calibration Burden: Challenges and Future. *Front. Bioeng. Biotechnol.* **2024**, *12*, 1329209. https://doi.org/10.3389/fbioe.2024.1329209.
65. Farrell, T. R.; Weir, R. F. The Optimal Controller Delay for Myoelectric Prostheses. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2007**, *15* (1), 111–118. https://doi.org/10.1109/TNSRE.2007.891391.
66. Bates, T. J.; Fergason, J. R.; Pierrie, S. N. Technological Advances in Prosthesis Design and Rehabilitation Following Upper Extremity Limb Loss. *Curr. Rev. Musculoskelet. Med.* **2020**, *13* (4), 485–493. https://doi.org/10.1007/s12178-020-09656-6.
67. Huinink, L. H. B.; Bouwsema, H.; Plettenburg, D. H.; Van Der Sluis, C. K.; Bongers, R. M. Learning to Use a Body-Powered Prosthesis: Changes in Functionality and Kinematics. *J. Neuro-*

Engineering Rehabil. **2016**, *13* (1), 90. https://doi.org/10.1186/s12984-016-0197-7.
68. Sharma, A.; Niu, W.; Hunt, C. L.; Levay, G.; Kaliki, R.; Thakor, N. V. Augmented Reality Prosthesis Training Setup for Motor Skill Enhancement. arXiv 2019. https://doi.org/10.48550/ARXIV.1903.01968.
69. Walch, G.; Boileau, P. Prosthetic Adaptability: A New Concept for Shoulder Arthroplasty. *J. Shoulder Elbow Surg.* **1999**, *8* (5), 443–451. https://doi.org/10.1016/S1058-2746(99)90074-5.
70. Raschke, S. U. Limb Prostheses: Industry 1.0 to 4.0: Perspectives on Technological Advances in Prosthetic Care. *Front. Rehabil. Sci.* **2022**, *3*, 854404. https://doi.org/10.3389/fresc.2022.854404.
71. Martarello, R. de A. Planned Obsolescence in Public Health Services: The Prosthetic Mafia Case. *Saúde E Soc.* **2022**, *31* (1), e200398. https://doi.org/10.1590/S0104-12902022200398.
72. McLarney, M.; Pezzin, L. E.; McGinley, E. L.; Prosser, L.; Dillingham, T. R. The Prevalence of Lower Limb Loss in Children and Associated Costs of Prosthetic Devices: A National Study of Commercial Insurance Claims. *Prosthet. Orthot. Int.* **2021**, *45* (2), 115–122. https://doi.org/10.1177/0309364620968645.
73. AT2030 Programme. *Product Narrative: Prostheses*; Global Disability Innovation Hub: London, UK, 2020. https://at2030.org/static/at2030_core/outputs/Prostheses_Product_Narrative_a11y_20200827.pdf (accessed 2025-12-08).
74. Manero, A.; Smith, P.; Sparkman, J.; Dombrowski, M.; Courbin, D.; Kester, A.; Womack, I.; Chi, A. Implementation of 3D Printing Technology in the Field of Prosthetics: Past, Present, and Future. *Int. J. Environ. Res. Public. Health* **2019**, *16* (9), 1641. https://doi.org/10.3390/ijerph16091641.
75. Thomas, D. J.; Singh, D. 3D Printing for Developing Patient Specific Cosmetic Prosthetics at the Point of Care. *Int. J. Surg.* **2020**, *80*, 241–242. https://doi.org/10.1016/j.ijsu.2020.04.023.
76. Jeong, E. C.; Lee, G. K. Abstract: The Use of a Low-Cost Three-Dimensional Printed Single Channel Myoelectric Interfaced Prosthesis for Transradial Amputees: Preliminary Report. *Plast. Reconstr. Surg. - Glob. Open* **2017**, *5* (9S), 147–148. https://doi.org/10.1097/01.GOX.0000526377.39346.9a.
77. Carro, R. M.; Costales, F. G.; Ortigosa, A. Serious Games for Training Myoelectric Prostheses through Multi-Contact Devices. *Children* **2022**, *9* (3), 423. https://doi.org/10.3390/children9030423.
78. Fietz, K.; Lay, J. Digital Skills in the Global South: Gaps, Needs, and Progress. *GIGA Focus Glob.* **2023**, *2*. https://doi.org/10.57671/GFGL-23022.
79. Gilotta, S.; Spada, S.; Ghibaudo, L.; Isoardi, M.; Mosso, C. O. Acceptability Beyond Usability: A Manufacturing Case Study. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*; Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, 2019; Vol. 824, pp 922–934. https://doi.org/10.1007/978-3-319-96071-5_95.
80. Össur. *Introduction to i-Limb.* Össur. https://www.ossur.com/en-ca/prosthetics/upper-limb-training/i-limb (accessed 2025-12-08).
81. Hammond, B.; Abu, K. A.; Anyittey-Kokor, D.; Baidoo, P. K.; Leat, M.; Awoonor-Williams, R.; Konadu-Yeboah, D.; Wilson, A. A.; Vormawor, K. K.; Bukari, M. I. S.; Twumasi-Baah Jnr, K.; Amamoo, J.; Opuni, G. Y. F.; Shiako, J. T.; Akomea, G. Determinants of Utilization of Prostheses and Orthoses Following Lower Limb Amputation in Sub-Saharan Africa: A Systematic Review. *J. Orthop. Rep.* **2025**, *4* (1), 100528. https://doi.org/10.1016/j.jorep.2024.100528.
82. Inoue, Y.; Yamanoi, Y.; Kuroda, Y.; Ono, Y.; Shimada, T.; Yabuki, Y.; Yokoi, H. Development of a Waterproof EMG Sensor for Improved Signal Quality by Using Load Balancing. *IEEE Sens. J.* **2025**, *25* (9), 14617–14628. https://doi.org/10.1109/JSEN.2025.3551507.
83. Bolivar, E.; Rezazadeh, S.; Summers, T.; Gregg, R. D. Robust Optimal Design of Energy Efficient Series Elastic Actuators: Application to a Powered Prosthetic Ankle. In *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*; IEEE: Toronto, ON, Canada, 2019; pp 740–747. https://doi.org/10.1109/ICORR.2019.8779446.
84. Dababneh, S.; Dababneh, N.; Xie, C.; Henchi, H.; Efanov, J. I. Shaping the Future of Upper Extremity Prostheses Through 3D Printing. *Prosthesis* **2025**, *7* (2), 39. https://doi.org/10.3390/prosthesis7020039.
85. Ribeiro, J.; Mota, F.; Cavalcante, T.; Nogueira, I.; Gondim, V.; Albuquerque, V.; Alexandria, A. Analysis of Man-Machine Interfaces in Upper-Limb Prosthesis: A Review. *Robotics* **2019**, *8* (1), 16. https://doi.org/10.3390/robotics8010016.
86. Liu, X.; Zhang, D.; Miao, K.; Guo, Y.; Jiang, X.; Zhang, X.; Jia, F.; Tang, H.; Dai, C. A Review on the Usability, Flexibility, Affinity, and Affordability of Virtual Technology for Rehabilitation Training of Upper Limb Amputees. *Bioengineering* **2023**, *10* (11), 1301. https://doi.org/10.3390/bioengineering10111301.
87. Jones, H.; Dupan, S.; Dyson, M.; Krasoulis, A.; Kenney, L. P. J.; Donovan-Hall, M.; Memarzadeh, K.; Day, S.; Coutinho, M.; Nazarpour, K. Co-Creation and User Perspectives for Upper Limb Prosthetics. *Front. Neurorobotics* **2021**, *15*, 689717. https://doi.org/10.3389/fnbot.2021.689717.
88. Coapt. Complete Control System Gen2 Handbook (V 5.1). https://coaptengineering.com/wp-content/uploads/2021/06/Coapt-Gen2-Handbook-v5.1.pdf (accessed 2026-06-06).
89. Luchetti, M.; Cutti, A. G.; Verni, G.; Sacchetti, R.; Rossi, N. Impact of Michelangelo Prosthetic Hand: Findings from a Crossover Longitudinal Study. *J. Rehabil. Res. Dev.* **2015**, *52* (5), 605–618. https://doi.org/10.1682/JRRD.2014.11.0283.
90. Fillauer. Hosmer 5 Hook. https://fillauer.com/product/hosmer-5-hook/ (accessed 2026-06-06).
91. Amputee Store. Hosmer Prosthetic Model 5X Hook, Stainless Steel. https://amputeestore.com/products/hosmer-model-5x-prosthetic-hook (accessed 2026-06-06).
92. Ottobock. Michelangelo Hand 8E500 Instructions for Use (Qualified Personnel), 2022. https://res.cloudinary.com/ottobock-se/image/upload/v1/Communication-medias/Technical%20communication%20medias/Instruction%20documents/ Manuals/Instructions%20for%20use%20(qualified%20personnel)/647G587/647G587-0%3DEN_INT%20-%20Instructions%20for%20use/Web/4817193?_a=BAMCkGiu0 (accessed 2026-06-06).
93. Ottobock. AxonSkin Natural. https://www.ottobock.com/en-us/b2b-product/8S502 (accessed 2026-06-06).
94. Loth Fabenim. Hosmer 5X, 5XA & 5XTi Hooks. https://www.lothfabenim.com/en/component/rsfiles/download-file/files?Itemid=4335&path=producten%252FArmprothesen%252FHaken%252FModel%2B5X%252FEnglish%2B-%2BLoth%2BFabenim%2BProduct%2BManual%2BHosmer%2B5X%252C%2B5XA%2B%2526%2B5XTi%2BHooks.pdf.

Authors

Sheraya Amarasekara, a student at Dubai College, is a student researcher with deep interests in affordable, user-focused prosthetic engineering and the use of genomics for cost-effective, preventative medicine in developing areas. She aspires to improve accessibility via medical policy while innovating as a biomedical engineer in the design and development sector.

DOI: 10.36838/IJHSR819.19

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Structure, Dysfunction, and Therapeutic Targeting of the p53 Tumor Suppressor in Human Cancer

Tanya Verma

Northview High School, 10625 Parsons Rd, Johns Creek, GA 30097, USA; verma.tanya1210@gmail.com

ABSTRACT: *TP53* is the most frequently mutated gene in human cancer, and its protein product, p53, serves as the cell's primary defense against cancer proliferation. Targeting p53 therapeutically has proven remarkably difficult, since p53 dysfunction cannot be pinpointed to a single problem. Therapies targeting p53 must be matched to specific dysfunction categories, such as loss of function, gain of function, or negative regulator overexpression. This review investigates how the diversity of p53 mutations necessitates targeted therapy by analyzing p53 structure, mutation types, and current therapeutic approaches, using evidence from experimental cancer models and clinical trials. By analyzing current treatment options, we show how p53 therapy relies on understanding mutation mechanisms. Structural mutants require reactivation drugs like APR-246 to refold the protein, while gain-of-function mutants require degradation strategies such as the Hsp90 inhibitor 17-AAG. Using the two interchangeably could render treatments ineffective or worsen outcomes by stabilizing oncogenic activity. This review argues that the effectiveness of p53-targeted therapy depends on specific, categorized treatment options. Matching the right therapy to the right mutation reduces treatment failure, improves clinical design, and could help save the lives of cancer patients with p53-altered tumors.

KEYWORDS: Cellular and Molecular Biology, Genetics; Cancer, Treatment, Therapeutic Strategies, Drugs, Tumor Suppression, Protein p53.

Introduction

The p53 protein is a central tumor suppressor that maintains genetic stability in most somatic cells and is commonly known to be a regulator of intracellular functions. Its loss of function due to a mutation in the *TP53* gene that allows for the proliferation of faulty DNA.[1,2] It is important to understand the structure and behavior of p53 when considering the development and progression of cancer, as p53 plays a crucial role in regulating the cell cycle in response to stress signals.[2] The presence of mutant p53 (mutp53) varies among cancer types, with certain cancers exhibiting higher mutation rates than others. Small cell lung cancer, colorectal cancer, ovarian cancer, and endometrial cancer all have mutation frequencies of over 50%. Genomic instability and increased resistance to therapies often occur as a result of high p53 mutation frequencies.[3] This can be shown through research that explores how high mutp53 expression in cancer patients is correlated with reduced overall patient survival.[3] Globally, approximately 20 million new cancer cases and 9.7 million cancer deaths were recorded in 2022 alone.[4] Given that TP53 mutations are present in over 50% of all human tumors, this translates to an estimated 10 million new p53-mutant cancer diagnoses annually.[4,5] These findings emphasize the importance of understanding the role and behavior of p53 when considering cancer development and progression.

Under normal conditions, p53 protein levels are kept under control by one of its primary negative regulators, MDM2. MDM2 and p53 form an autoregulatory feedback loop that targets p53 for degradation, ensuring activity remains low in healthy cells.[6] Upon activation, p53 induces the expression of genes such as p21 that mediate p53-induced G1 cell cycle arrest by inhibiting cyclin-dependent kinases (CDKs). This prevents RB phosphorylation, which halts entry into the S phase. As a result, cells are given more time to repair damaged DNA or undergo apoptosis.[7] Additional studies have shown that p53 controls many "non-classical" pathways, including metabolic homeostasis, senescence, stem cell differentiation, angiogenesis, and DNA repair. Wild-type p53 (wtp53) also influences metabolic processes through transcriptional activation or by repressing target genes. p53 can suppress the expression of glucose transporters like GLUT1 and GLUT4 to reduce glucose uptake, which protects cells from oxidative stress. p53 promotes mitochondrial respiration by upregulating genes that promote oxidative phosphorylation, electron transport chain activity, and efficient ATP production. By increasing the expression of genes that promote fatty acid oxidation (FAO), p53 helps fatty acids enter the mitochondria, where they are broken down to produce ATP.[6,8]

This paper will discuss how mutations in p53, such as regulatory expression by negative modulators, aberrant protein-protein interactions, and structural missense mutations, can contribute to cancer development and progression. We will go over how such mutations can be therapeutically targeted through different strategies such as reactivation therapies, inhibition of MDM2 and MDMX (also known as MDM4), and deletion or degradation of mutp53 proteins. The main therapeutic focus will be on small-molecule-based therapies that aim to reactivate mutp53 to a functional state and stabilize wtp53. Despite extensive research, p53 remains one of the most difficult tumor suppressor proteins to target.[1] Recent research shows that treating p53 dysfunctions the same regardless of the specific mutation category is not effective, nor is it accurate.

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.20

Different types of mutations in p53 lead to different cancer behaviors that should not be treated interchangeably, which raises the issue of certain therapies failing as they are unable to account for the specific p53 mutation situation. Recent studies in biology and pathway analysis show there is an urgent need to distinguish between loss-of-function (LOF) and gain-of-function (GOF) mutations and to identify when p53 should be neutralized or reactivated.[9] Understanding and addressing these points is crucial for developing new therapies and advancing oncology for p53-driven cancers in general.

Discussion

Structure and Normal Function of p53:

p53 functions as a sequence-specific transcription factor that uses stress signals to indicate issues such as DNA damage, oncogene activation, ribosomal stress, and nutrient deprivation. In response to these signals, it regulates genes that control whether the cell repairs damage, stops division, or undergoes apoptosis.[10] In humans, the *TP53* gene is located on the short arm of chromosome 17 at position 13.1, and encodes p53 with 393 amino acid residues. The 393 amino acids that p53 consists of make up its domains, which enable tumor-suppressing functions.[11] p53 functions as a homotetramer because individual monomers bind DNA weakly; thus, tetramer formation is essential for sequence-specific DNA binding.[12] Studies show that the tetramerization domain (TD) in the C-terminal region and the DNA binding domain (DBD) act together to make p53 a functional transcription factor capable of regulating genes.[11] Tetramer formation is essential for DNA binding and for ensuring p53 can be properly regulated to effectively activate its target genes in response to stress.[13,14] Furthermore, studies show how minor disruptions in tetramer formation can severely impair gene regulation.[15]

The p53 protein relies mainly on domains for its tumor-suppressing functions. These domains allow for the binding of p53 to other cofactors, also known as "helper proteins," that are crucial for p53 to be able to activate or suppress genes. When these domains are missing or mutated, p53's tumor-suppressing function is weakened, which can lead to cancer formation and growth.[16] The N-terminal region of p53 is primarily responsible for activating target gene expression. This region contains two transcriptional activation (transactivation) domains located within residues 1–61. TAD1 (residues 1–40) and TAD2 (residues 40–61) both help initiate the transcription process.[17,18] TADs also interact with one of p53's negative regulators, MDM2, an E3 ubiquitin ligase that marks p53 for degradation under normal conditions, ensuring p53 levels stay low when cells are not stressed.[19] Additionally, the DNA-binding domain makes connections with both TADs by folding back and contacting parts of the same p53 molecule. These contacts help p53 identify target DNA sequences.[20]

The proline-rich domain (PRD, residues ~63-94) has a unique flexible structure that helps stabilize p53 by regulating interactions with proteins such as MDM2. Through studies, it has been demonstrated that the PRD influences how p53 binds to DNA and how effectively it activates or represses different target genes. Unlike the DNA-binding domain, which is frequently mutated in cancer, the PRD is rarely mutated.[21] The central DNA-binding domain (DBD, residues ~102-293) is the most critical region for p53 activity as a transcription factor and tumor suppressor. The DBD is organized around an antiparallel β-sheet scaffold and contains several flexible loop regions, including the L1, L2, and L3 loops, which help with DNA recognition. These loop regions make sequence-specific contacts with DNA by inserting into the major and minor grooves of the DNA helix.[22] The DBD functions as part of a tetrameric complex, with four p53 monomers assembling to bind DNA cooperatively (Figure 1). The three-dimensional fold of the DBD is maintained by a zinc ion, which is coordinated by conserved cysteine and histidine residues.[23] Most cancer-associated mutations occur in the DBD, with over 80% of p53 mutations found in tumors located in this domain.[22]

The C-terminal regulatory domain (CTD, residues 363-393) modulates how, when, and how strongly p53 binds DNA and activates target genes. Flexibility and lack of three-dimensional structure in the CTD enable it to interact with multiple binding partners and respond better to regulatory signals.[24] It is one of the p53 domains that is subject to extensive post-translational modifications, including phosphorylation, acetylation, and ubiquitination.[25] These modifications change p53's interaction with DNA and are crucial for determining whether p53's ability to bind to target sequences will be enhanced or inhibited. Ensuring all five domains work together coordinatedly is essential for ensuring p53 does its job as a tumor suppressor effectively, with each domain contributing differently to how p53 senses stress, binds DNA, and activates gene expression programs.

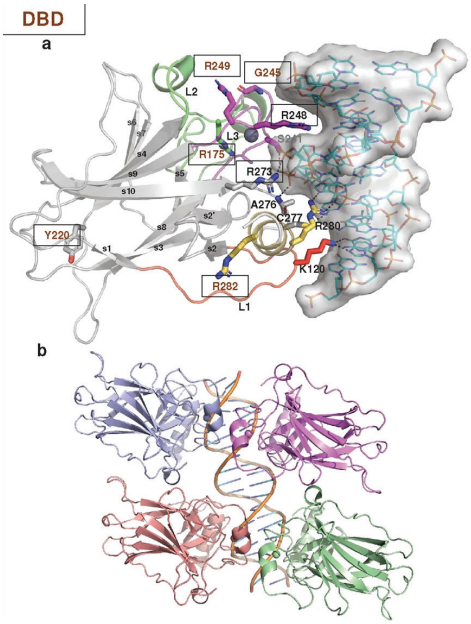


Figure 1: Structural organization of the p53 DNA-binding domain (DBD) and tetramer. a) three-dimensional structure of the DBD in complex with sequence-specific DNA, highlighting the β-sheet scaffold and loop regions (L1, L2, L3) that make direct contact with DNA bases. b) Full p53 tetramer bound to DNA, illustrating how four monomers assemble into two dimers that each recognize one DNA half-site. This tetrameric arrangement is essential for p53's function as a transcription factor. Minor disruptions to this structure, such as those caused by structural hotspot mutations, can abolish tumor suppressor activity. Reproduced from Wang *et al.* 2023 under CC BY 4.0 license.[1]

MDM2-mediated ubiquitination and proteasomal degradation are two post-translational mechanisms that work together to unfold and degrade the protein, keeping p53 levels low.[26] Upon cellular stress signals such as DNA damage, oncogene activation, or hypoxia, p53 undergoes post-translational modifications that disrupt the p53-MDM2 interaction. Stress-activated kinases, including ATM, ATR, and CHK1/CHK2, phosphorylate p53 at Ser15 and Ser20, disrupting the p53-MDM2 interaction and allowing p53 to accumulate.[27,28] These modifications allow p53 to accumulate rapidly and form tetramers in the cell. Once in the nucleus, p53 must locate and bind to DNA-binding sites as a tetramer.[12]

Once stabilized, p53 activates target genes appropriate to the stress signal. For cell-cycle arrest, p53 strongly induces CDKN1A, a cyclin-dependent kinase inhibitor that codes for p21 and blocks cell cycle progression at the G1/S checkpoint. This halts the DNA replication process and gives cells time to repair damage before dividing.[7] In the case of irreparable damage, p53 activates pro-apoptotic genes such as BAX and PUMA, which trigger mitochondrial membrane permeabilization and caspase-mediated apoptosis.[29] For DNA repair, p53 induces genes like GADD45A, which facilitate repair processes and assist in chromatin remodeling during repair. p53 also induces expression of DNA damage recognition proteins DDB2 and XPC, which recognize DNA lesions and help recruit repair machinery to sites of damage.[30] The structural and physiological integrity of p53 is crucial for it to carry out its functions. The DNA-binding domain must be properly positioned and structured to enable the recognition of target DNA sequences. Tetramer formation through the tetramerization domain provides the stability and structure required for proper DNA binding. Both the N-terminal and C-terminal regions allow these domains to integrate regulatory signals that fine-tune p53's activity (Figure 2).

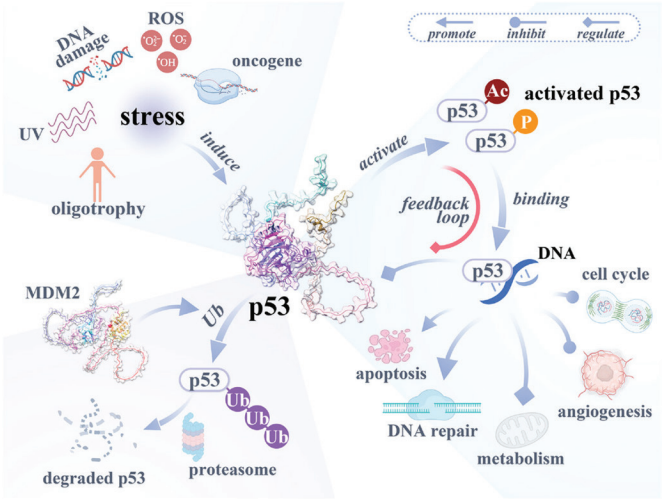


Figure 2: The p53 signaling pathway under normal and stress conditions. Under normal conditions, MDM2 ubiquitinates p53, targeting it for proteasomal degradation through a negative feedback loop. Upon stress signals, including DNA damage, ROS, UV radiation, and oncogene activation, p53 is stabilized through phosphorylation and acetylation. This allows it to bind DNA and activate programs including cell cycle arrest, apoptosis, DNA repair, angiogenesis, and metabolic regulation. Reproduced from Shen *et al.* 2023 under CC BY 4.0 license.[31]

Over 80% of cancer-associated p53 missense mutations occur within the DBD, primarily clustered at specific residue positions known as mutation hotspots.[3] The six most frequently mutated residues are R175, G245, R248, R249, R273, and R282. These six residues account for a disproportionate amount of all p53 mutations across human cancers, as altering them disrupts p53 function, which gives cancer cells a survival advantage.[32] These hotspot mutations can be classified into two major categories based on their mechanisms of inactivation. Contact mutations, which include R248 and R273, occur at residues that make direct physical contact with DNA bases or the DNA backbone. Mutations at these locations eliminate hydrogen bonds and electrostatic interactions that enable p53's sequence-specific recognition abilities.[33,34] Structural mutations, which include R175H, G245S, R249S, and R282W, disrupt and destabilize the three-dimensional fold of the DBD, causing conformational changes or partial unfolding of the domain.[3,33] R175, for instance, disrupts zinc-binding site interactions and impairs loop positioning critical for DNA recognition. It is important to note that not all mutations at the same residue have identical effects; some mutations can cause complete loss of function, while others retain partial activity.[3] The high frequency of mutations specifically at these hotspots demonstrates the importance of these residues for p53's tumor suppressor activity and how their disruption promotes cancer progression.

p53 Mutations and Misregulation in Cancer:

Most *TP53* mutations are missense mutations, with over 80% involving single amino acid substitutions. This produces full-length proteins that are structurally altered instead of eliminating protein expression.[3,35] The majority of these missense mutations cluster within the DNA-binding domain between codons 125 and 300.[36] The consequences of p53 mutations can be classified based on their functional effects. Mutations can lead to loss-of-function, where the protein loses its normal tumor-suppressing activities, or gain-of-function, where the mutant protein acquires new properties that actively promote cancer progression.[35,37] Because p53 functions as a tetramer, mutp53 proteins can exert dominant-negative effects by forming mixed tetramers with remaining wtp53 and blocking its function.[38]

Loss-of-function mutations are the most common form of p53 inactivation in cancer. When p53 loses the ability to activate its target genes, cells lose access to the stress-response programs that would otherwise protect them from accumulating genetic damage. As a result, cells are no longer able to carry out processes such as inducing cell cycle arrest through p21, triggering apoptosis through BAX and PUMA, or repairing damaged DNA through genes like GADD45A, DDB2, and XPC.[7,29,30] Without functional pathways, additional mutations can arise and go uncorrected. Loss-of-function mutations actively create conditions where genetically damaged cells continue to survive and replicate, further enabling cancer development.[3]

Unlike LOF mutations, which eliminate p53's normal tumor-suppressing abilities, GOF mutations give mutant p53

DOI: 10.36838/IJHSR819.20

new cancer-promoting abilities. GOF mutp53 remains structurally stable and begins interacting abnormally with other proteins in the cell.[35,38] One of the most commonly observed GOF mechanisms involves mutp53 binding to and blocking p63 and p73, two proteins from the same family as p53 that share a similar structure and function. By interacting with them, mutp53 prevents p63 and p73 from binding to their DNA targets, which blocks their ability to activate genes that suppress cell growth and trigger apoptosis.[2,39] Gain of function mutp53 also binds with other transcription factors such as NF-κB, ETS2, NF-Y, SREBP, and HIF-1α. The mutp53 then changes which genes get turned on or off in ways that benefit cancer cells.[2] GOF mutp53 is also able to enhance the ability of cancer cells to invade surrounding tissue and spread to other parts of the body by turning on genes involved in cell movement and tissue breakdown.[38] It increases resistance to chemotherapy by activating proteins that either pump drugs out of cancer cells or break them down before they can work.[38] Additionally, by interfering with DNA repair and cell division checkpoints, GOF mutp53 can increase cell duplication and contribute to DNA damage.[38]

The differences between LOF and GOF mutations make it clear that not all p53 mutations work the same way (Figure 3). Loss-of-function mutations prevent p53 from binding to DNA and activating the genes that protect against cancer. Typically, the protein will misfold, lose its ability to bind to DNA, or both. Gain-of-function mutations work differently. The mutant p53 will no longer be able to function as a normal transcription factor, but instead of becoming inactive, it begins interacting with other proteins in harmful ways.[35,39] The two types also differ in protein stability. Loss-of-function mutants are often structurally unstable and may break down or co-aggregate inside cells. Gain-of-function mutants are typically very stable. They accumulate to abnormally high levels in cancer cells, often because chaperone proteins like HSP90 protect them from degradation.[2] Clinically, tumors with GOF mutations tend to be more aggressive, harder to treat, and associated with worse patient outcomes.[35] While LOF mutations eliminate p53's tumor-suppressive capacity, GOF mutations give the protein new oncogenic properties that contribute to tumor progression. Because of their prominent differences, making the distinction between these two mutation types is critical for understanding and treating p53-driven cancers.

Mutant p53 proteins can also cause additional damage by mixing with wtp53 tetramer subunits and disrupting their function, also known as the dominant-negative effect. In cells that carry one mutant and one normal copy of *TP53*, both types of p53 protein are produced and can combine into the same tetramer.[40] This means that even one mutant copy of p53 can impair and affect the function of the remaining wtp53. Studies have shown that cells with equal amounts of mutant and wild-type p53 display less than 50% of normal p53 DNA-binding activity. This further suggests that mutp53 dominant negative effects are disproportionately damaging.[41] This has been confirmed in animal studies where mice carrying one mutant *TP53* gene and one normal *TP53* gene behave similarly to mice completely lacking p53 when exposed to DNA-damaging agents.[40,42] This helps explain why the Li-Fraumeni syndrome causes cancers at very high rates and unusually early ages, even though they still carry one functional p53 allele.[41] It is important to understand that the dominant-negative effect is not equally strong for all mutations and varies depending on the specific mutation or cell type. Over 91% of p53-mutant cancers ultimately lose both p53 gene alleles, which indicates evolutionary pressure on the cancer cells to fully remove p53 function rather than relying on partial removal through the dominant-negative effect.[43]

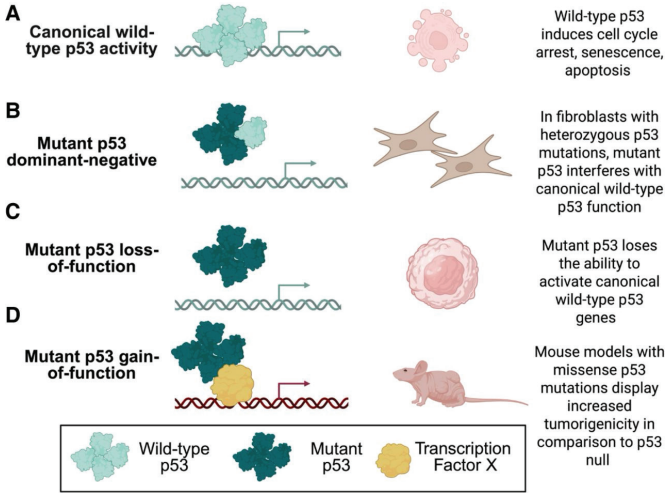


Figure 3: Functional categories of mutant p53 and their mechanisms of tumor suppressor inactivation. Wild-type p53 activates target genes that suppress cancer growth. Loss-of-function (LOF) mutations prevent p53 from binding DNA, eliminating its tumor-suppressive capacity. Gain-of-function (GOF) mutations give the mutant protein new cancer-promoting interactions with transcription factors. The dominant-negative effect occurs when mutant and wild-type p53 combine into mixed tetramers, impairing the remaining functional protein. Each category represents a fundamentally different molecular problem requiring a distinct therapeutic strategy. Reproduced from Efe *et al.* 2026 under CC BY 4.0 license.[37]

p53 can also be functionally silenced in cells with a completely normal, unmutated gene. One of the most common ways this happens is through the overexpression of MDM2 or its related protein MDMX. In normal cells, MDM2 and p53 exist in a balanced feedback loop that keeps p53 levels low when cells are healthy. In certain cancers, however, the genes encoding MDM2 or MDMX are amplified or overexpressed. As a result, these proteins are produced in excess and continuously suppress p53 activity even when the cell is under stress.[44,45] High-risk human papillomaviruses like HPV-16 and HPV-18 produce a protein called E6 that directly degrades p53 through a pathway distinct from MDM2, achieving the same functional effect as a mutation.[46,47] p53 can also be impaired when the cellular machinery responsible for activating it is disrupted. For example, if the kinases that phosphorylate p53 at S15 and S20 in response to DNA damage are lost or mutated, p53 cannot be properly activated. This leaves cells functionally p53-deficient despite having a completely intact *TP53* gene.[44] These non-mutational mechanisms demonstrate that p53 dysfunction can occur through multiple pathways, each of which may require a different therapeutic approach.

Therapeutic Targeting of p53 Alterations:

p53 is one of the most common targets in cancer therapy, given its tumor-suppressing role in the cell cycle and the high frequency of its dysfunction in human cancers. However, due to its lack of typical drug target features such as well-defined binding pockets, p53 was considered an "undruggable" protein for many years.[48-50] The challenge is further complicated by the fact that p53 dysfunction in cancer does not follow a single pattern. Some tumors carry LOF mutations, while others carry GOF mutations. wtp53 is also commonly suppressed by regulatory mechanisms like MDM2 overexpression.[51] Each scenario presents a different problem that cannot be addressed by the same therapeutic approach. As a result, p53 targeted therapies can be sorted into different categories. The most current major approaches include reactivation of misfolded mutp53 proteins, inhibition of negative regulators like MDM2 and MDMX, degradation of GOF mutp53 proteins, and small-molecule strategies that can be tailored to specific mutation contexts (Table 1).[1]

Overexpression of MDM2 or MDMX presents the therapeutic goal of releasing and stabilizing the functional p53 that is already present, since this type of dysfunction cannot be directly categorized as a mutation. MDM2 suppresses p53 by blocking target gene activation, marking p53 for degradation, and exporting p53 out of the nucleus.[52] However, MDMX works differently. Instead of degrading p53 directly, it binds to the same transactivation domain as MDM2 and suppresses p53's transcriptional activity. MDM2 and MDMX form heterodimers to increase the efficiency of p53 ubiquitination and degradation.[44] Small molecules serve as a therapeutic solution for this by binding to the p53-binding pocket of MDM2, which physically blocks it from interacting with p53. The most well-studied examples are the Nutlin family of compounds, including Nutlin-3 and its clinical derivatives RG7112 and idasanutlin (RG7388). These small molecules occupy MDM2's hydrophobic binding cleft by mimicking three of p53's key binding residues (Phe19, Trp23, and Leu26). MDM2 is blocked from suppressing p53 as a result, allowing p53 levels to accumulate and its function to be restored.[53]

Despite their initial clinical promise, MDM2 inhibitors have not achieved widespread therapeutic success. This is due to many factors, such as toxicity, side effects that limit patient tolerance, and rapid development of resistance through acquired p53 mutations. A phase I clinical trial of RG7112 (a first-generation Nutlin compound) showed responses in AML patients, but all patients experienced adverse events, including dose-limiting thrombocytopenia, neutropenia, diarrhea, and gastrointestinal toxicity.[54] Another RG7112 trial in 20 patients with MDM2-amplified liposarcoma demonstrated 70% stable disease with increased p53 expression, but late hematological toxicity remained a significant concern.[55] Resistance studies revealed that prolonged treatment with idasanutlin (RG7388, a second-generation MDM2 inhibitor) led to the emergence of p53-mutated resistant cell populations, demonstrating that these inhibitors provide limited cancer cell elimination and select for p53 LOF mutations.[56] Additionally, MDM2 inhibitors cannot fully activate p53 in tumors where MDMX is also overexpressed. The study of dual MDM2/MDMX inhibitors and combination strategies that target both proteins simultaneously has increased in popularity as a result.57 These approaches are only effective in tumors where p53 remains structurally intact.

For tumors carrying missense p53 mutations, the therapeutic goal involves fixing the protein directly. Many missense mutations, particularly structural ones like R175H and R249S, cause the DNA-binding domain to lose its proper shape without destroying the protein entirely. The protein is still there and stable, but it is misfolded, which prevents DNA binding. If a small molecule can restore the protein back into its proper shape, it might regain the ability to bind DNA.[48] The most advanced drug in this category is APR-246 (also known as eprenetapopt or PRIMA-1MET), which is a methylated derivative of the earlier compound PRIMA-1. Inside cells, APR-246 converts to an active compound called methylene quinuclidinone (MQ). MQ works by forming covalent bonds with cysteine residues on the p53 surface, particularly Cys124 and Cys277. These bonds stabilize the mutant protein and shift its shape back toward a wild-type-like structure that can bind DNA and turn on target genes.[1]

Early clinical trials demonstrated remarkable promise in patients with *TP53*-mutant myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML). In a phase Ib/II trial, APR-246 was combined with azacitidine, the first FDA-approved treatment of MDS. The combination achieved an overall response rate of 71% with 44% complete remission (CR) in 55 patients with *TP53*-mutant MDS or AML.[58] A similar French phase II study confirmed these results, reporting a 62% overall response rate in MDS patients with a median overall survival of 12.1 months.[59] The subsequent phase III trial, however, did not provide statistical significance even though the complete remission rate was 53% higher in the eprenetapopt plus azacitidine arm (33.3%) compared to azacitidine alone (22.4%).[60] Another phase II trial evaluating APR-246 plus azacitidine as post-transplant maintenance therapy in *TP53*-mutant MDS and AML patients achieved a median relapse-free survival of 12.5 months and median overall survival of 20.6 months, a substantial improvement compared to historical controls showing approximately 30% one-year relapse-free survival and median overall survival of 5-8 months after transplantation.[61]

A related approach uses zinc metallochaperones. Structural mutations such as R175H prevent the mutant protein from properly coordinating the zinc ion that is essential for correct DNA-binding domain folding. Therapeutic Zinc metallochaperones like ZMC1 (NSC319726) are synthetic molecules that transport zinc into the cell, allowing it to bind to the zinc-binding site of mutp53, restoring zinc coordination and allowing the protein to refold into its normal functional shape.[62] These results demonstrate that mutp53 reactivation is achievable through stabilization approaches like APR-246 and zinc metallochaperone strategies like ZMC1. Ongoing trials continue to evaluate these therapies in different contexts and new strategies for targeting missense mutations.

DOI: 10.36838/IJHSR819.20

Another reactivation strategy involves COTI-2, an orally available third-generation thiosemicarbazone compound. The drug reshapes mutp53 back into wtp53 while also independently triggering DNA damage responses that push cancer cells toward apoptosis.[63] This means COTI-2 can eliminate tumor cells through both p53-dependent and p53-independent pathways, which is a significant advantage in tumors where restoring p53 function alone may not be enough to stop cancer cell survival.[63] Preclinical studies in head and neck cancer showed that COTI-2 also increases the effectiveness of cisplatin and radiation when used together, suggesting potential combined therapeutic approaches.[63] A Phase I clinical trial evaluating COTI-2 in patients with recurrent gynecologic cancers found it to be generally safe and well tolerated, with the most common side effects being nausea, fatigue, and gastrointestinal symptoms.64 However, like other therapeutic approaches, COTI-2 is not a universal solution, and its effectiveness varies depending on the specific mutation present.[1,63]

Under normal conditions, p53 must be present in the nucleus to carry out its tumor-suppressing functions, but a protein called XPO1 acts as a nuclear export receptor that physically transports p53 out of the nucleus and into the cytoplasm.[65] In cancer, XPO1 is frequently overexpressed, causing excessive export of p53 and other tumor suppressor proteins away from their targets, contributing directly to cancer progression.[65] This problem is compounded by MDM2, which not only marks p53 for degradation but also actively facilitates its nuclear export.[26] Nuclear targeting strategies address this by blocking XPO1 directly, trapping p53 inside the nucleus where it can function.[66] The most clinically advanced drug in this category is Selinexor (KPT-330), an orally available XPO1 inhibitor. Studies in gastric cancer found that XPO1 was overexpressed in tumor tissue compared to normal tissue and was associated with poor patient survival, and that blocking XPO1 with Selinexor effectively caused p53 to accumulate in the nucleus.[66] Selinexor also provides opportunities for combination approaches, with greater anti-tumor effects shown when combined with standard chemotherapy compared to either treatment alone.[66] Nuclear targeting is particularly relevant for tumors where p53 is structurally intact but kept inactive through regulatory suppression, making it a complementary approach to other reactivation drugs like APR-246, ZMC1, and COTI-2.[1,65,66]

Gain-of-function mutp53 presents a different problem. Restoring wild-type shape through drugs like APR-246 will most likely be of no help here because the main problem is not the absence of tumor suppression, but the promotion of cancer through aberrant protein interactions.[67] One major strategy for GOF mutants is promoting degradation. In cancer cells, chaperone proteins, including Hsp40, Hsp70, and Hsp90, bind to and shield mutant p53 from the normal ubiquitin-proteasome degradation machinery.[67] Disrupting these interactions will re-expose mutp53 to degradation, reducing its levels and eliminating its oncogenic activity. The Hsp90 inhibitor 17-AAG (tanespimycin) has been shown to destabilize mutp53 in cancer cells, with preclinical studies that demonstrate reduced mutp53 levels and tumor growth inhibition.[67] Combining it with MDM2 inhibitors like nutlin-3a produces synergistic effects that destabilize mutp53 while activating any remaining wtp53.[44] However, this combination has not yet been tested in clinical trials.

The chaperone-associated E3 ubiquitin ligase CHIP (the C-terminus of Hsc70-interacting protein) is also a possible solution for targeting mutp53 for degradation. It recognizes misfolded proteins such as mutp53 and directly ubiquitinates them for proteasomal degradation. CHIP activity is not transcriptionally regulated by p53, meaning it can function even when the p53-MDM2 feedback loop is broken. While this could prove to be an effective therapy method, direct CHIP activators do not currently exist, and this approach remains in early preclinical stages.[67] Another emerging strategy involves disrupting the specific GOF mutant protein-protein interactions. Mutant p53 binding to transcription factors such as p63, p73, and NF-κB enables GOF mutant activity. Small molecules that interfere with these interactions could potentially block GOF activity without having to restore wild-type function or degrade the protein. This approach is still in earlier stages of development compared to other therapies, but it presents a more targeted way to address specifically GOF mutations.[67]

Table 1: Mutation-specific therapeutic strategies targeting p53 dysfunction. Structural mutants are targeted through reactivation drugs like APR-246 and zinc metallochaperones like ZMC1. COTI-2 offers an additional reactivation approach with a distinct mechanism. GOF mutants are addressed through degradation strategies, including 17-AAG and the CHIP pathway. Wild-type p53 suppressed by MDM2/MDMX is released through inhibitors like Nutlin-3a. Selinexor addresses nuclear export of p53 through XPO1 inhibition. Each strategy is matched to a specific type of p53 dysfunction, as applying the wrong approach to the wrong mutation context can worsen outcomes rather than improve them.

Method/Approach	Example Drugs	Clinical Status
Reactivation of structural mutant p53 to restore wild-type conformation	APR-246 (eprenetapopt), PRIMA-1MET	Phase III (MDS/AML)
Restoration of zinc coordination to structural mutants	ZMC1 (NSC319726)	Preclinical
Degradation of GOF mutant p53 using HSP90 inhibitors	17-AAG (tanespimycin)	Preclinical
Prevention of wild-type p53 degradation by blocking MDM2/MDMX	Nutlin-3a, RG7112	Phase I/II
E3 ubiquitin ligase-mediated degradation (CHIP pathway)	No direct activators available	Early preclinical
Reactivation of mutant p53 plus induction of DNA damage responses	COTI-2	Phase I
Nuclear retention/reactivation of p53 through XPO1 inhibition	Selinexor (KPT-330)	Phase I/II in p53-related solid tumors

p53 alterations are so diverse that effective therapies can only be developed if they are curated for specific types of dysfunction. Structural mutants like R175H that cause the DNA-binding domain to misfold are good candidates for reactivation strategies like APR-246, COTI-2, and zinc metallochaperones. Since these mutations disrupt protein structure rather than eliminating DNA-contacting residues, refolding the protein back into its proper shape can restore its normal abilities.[68] GOF mutants are best addressed through degradation strategies or disruption of their protein-protein interactions rather than reactivation. Simply trying to fix the protein may not eliminate its ability to hinder remaining healthy p53 or interact with other cancer-promoting proteins.[69] In contrast, tumors with wtp53 that are being suppressed through MDM2 or MDMX overexpression do not require structural repair. They instead

need MDM2/MDMX inhibitors to release p53 from suppression.[70] Tumors with intact p53 may also benefit from nuclear retention strategies such as XPO1 inhibition, which keeps p53 in the nucleus as a transcription factor.[66] Choosing the right therapeutic strategy based on the mutation context matters because using the wrong one can backfire. For example, using a reactivation drug like APR-246 on a GOF mutant may not be effective, since the primary problem is not misfolding but the mutant protein's aberrant interactions with transcription factors like p63 and p73, making degradation strategies like 17-AAG a more appropriate choice. Similarly, using MDM2 inhibitors in tumors with mutant p53 would stabilize the mutant protein instead of wild-type p53, potentially accelerating tumor growth.[57]

Conclusion

The p53 protein is essential for maintaining genetic stability through cell cycle regulation. As this review has demonstrated, p53's function depends on its structural integrity. The distinction between LOF and GOF mutations is critical to maintain this. Loss-of-function mutations eliminate tumor-suppressing activity while gain-of-function mutations drive cancer progression through new oncogenic properties. Beyond mutations, the overexpression of MDM2 and the dominant negative effect can further complicate treatment.

Promising mutation-specific therapies have begun to emerge. Structural mutants can be treated through drugs like APR-246, which aim to reactivate mutp53 by binding to specific cysteine residues and restoring its wild-type conformation. GOF mutants require degradation strategies rather than reactivation, since only refolding will not eliminate oncogenic activity. Wild-type p53 suppressed by MDM2 needs inhibitors to release it from suppression. Using the wrong approach can compromise clinical efficacy. For example, MDM2 inhibitors in mutp53 tumors can inadvertently stabilize the mutant protein, potentially worsening outcomes.

While mutation-specific treatment is increasingly practiced in clinical oncology, continued progress depends on expanding the range of effective drugs available for each mutation category. This requires several key advances. First, while molecular testing is increasingly used in clinical trials, broader and more consistent implementation across all p53-driven cancer types remains an important priority to ensure that drug mechanisms are matched to specific mutations. Second, combination therapies should be an area of exploration to address multiple aspects, such as the primary mutation and resistance pathways. Finally, new drugs should be developed for mutation types that currently lack effective treatments, particularly contact mutants and GOF degradation strategies. Matching therapies to specific p53 mutation types is key to improving cancer treatment.

Acknowledgments

I would like to acknowledge and thank my research mentors, Drs. Shaye and Tetz, for their guidance, support, subject knowledge, and feedback, which greatly contributed to the development of this study, and their encouragement pushed me to refine my work. I attest that the ideas, graphics, and writing in this paper are entirely my own.

References

1. Wang, H., Guo, M., Wei, H. *et al.* Targeting p53 pathways: mechanisms, structures and advances in therapy. Sig Transduct Target Ther 8, 92 (2023). https://doi.org/10.1038/s41392-023-01347-1
2. Alvarado-Ortiz, E., de la Cruz-López, K. G., Becerril-Rico, J., Sarabia-Sánchez, M. A., Ortiz-Sánchez, E., & García-Carrancá, A. (2021). Mutant p53 Gain-of-Function: Role in Cancer Development, Progression, and Therapeutic Approaches. Frontiers in cell and developmental biology, 8, 607670. https://doi.org/10.3389/fcell.2020.607670
3. Chen, X., Zhang, T., Su, W. *et al.* Mutant p53 in cancer: from molecular mechanism to therapeutic modulation. Cell Death Dis 13, 974 (2022). https://doi.org/10.1038/s41419-022-05408-1
4. Bray F, Laversanne M, Sung H, *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263. doi:10.3322/caac.21834
5. Wang C, Tan JYM, Chitkara N, Bhatt S. TP53 Mutation-Mediated Immune Evasion in Cancer: Mechanisms and Therapeutic Implications. Cancers (Basel). 2024 Sep 3;16(17):3069. doi: 10.3390/cancers16173069. PMID: 39272927; PMCID: PMC11393945.
6. Matthieu Lacroix, Romain Riscal, Giuseppe Arena, Laetitia Karine Linares, Laurent Le Cam, Metabolic functions of the tumor suppressor p53: Implications in normal physiology, metabolic disorders, and cancer, Molecular Metabolism, Volume 33, 2020, Pages 2-22, ISSN 2212-8778, https://doi.org/10.1016/j.molmet.2019.10.002
7. Engeland, K. Cell cycle regulation: p53-p21-RB signaling. Cell Death Differ 29, 946–960 (2022). https://doi.org/10.1038/s41418-022-00988-z
8. Flöter, J., Kaymak, I., & Schulze, A. (2017). Regulation of Metabolic Activity by p53. Metabolites, 7(2), 21. https://doi.org/10.3390/metabo7020021
9. Tornesello, M.L. (2025). TP53 mutations in cancer: Molecular features and therapeutic opportunities (Review). International Journal of Molecular Medicine, 55, 7. https://doi.org/10.3892/ijmm.2024.5448
10. Understanding the complexity of p53 in a new era of tumor suppression, Cancer Cell, Volume 42, Issue 6, 2024, Pages 946-967, ISSN 1535-6108, https://doi.org/10.1016/j.ccell.2024.04.009
11. Masahiro Nishimura, Yoshimasa Takizawa, Kayo Nozawa, Hitoshi Kurumizaka, Structural basis for p53 binding to its nucleosomal target DNA sequence, PNAS Nexus, Volume 1, Issue 4, September 2022, pgac177, https://doi.org/10.1093/pnasnexus/pgac177
12. Structural Basis of DNA Recognition by p53 Tetramers, Molecular Cell, Volume 22, Issue 6, 2006, Pages 741-753, ISSN 1097-2765, https://doi.org/10.1016/j.molcel.2006.05.015
13. McLure, K.G., Lee, P.W. How p53 binds DNA as a tetramer. EMBO J 17, 3342–3350 (1998). https://doi.org/10.1093/emboj/17.12.3342
14. Chène, P. The role of tetramerization in p53 function. Oncogene 20, 2611–2617 (2001). https://doi.org/10.1038/sj.onc.1204373
15. Rui Kamada, Takao Nomura, Carl W. Anderson, Kazuyasu Sakaguchi, Cancer-associated p53 Tetramerization Domain Mutants: QUANTITATIVE ANALYSIS REVEALS A LOW THRESHOLD FOR TUMOR SUPPRESSOR INACTIVATION, Journal of Biological Chemistry, Volume 286, Issue 1, 2011, Pages 252-258, ISSN 0021-9258, https://doi.org/10.1074/jbc.M110.174698
16. Liz J. Hernández Borrero, Wafik S. El-Deiry, Tumor suppressor p53: Biology, signaling pathways, and therapeutic targeting,

DOI: 10.36838/IJHSR819.20

Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, Volume 1876, Issue 1, 2021, 188556, ISSN 0304-419X, https://doi.org/10.1016/j.bbcan.2021.188556
17. Sullivan, K., Galbraith, M., Andrysik, Z. *et al.* Mechanisms of transcriptional regulation by p53. Cell Death Differ 25, 133–143 (2018). https://doi.org/10.1038/cdd.2017.174
18. Raj, N., & Attardi, L. D. (2017). The Transactivation Domains of the p53 Protein. Cold Spring Harbor perspectives in medicine, 7(1), a026047. https://doi.org/10.1101/cshperspect.a026047
19. Shan, B., Li, D. W., Brüschweiler-Li, L., & Brüschweiler, R. (2012). Competitive binding between dynamic p53 transactivation subdomains to human MDM2 protein: implications for regulating the p53·MDM2/MDMX interaction. The Journal of Biological Chemistry, 287(36), 30376–30384. https://doi.org/10.1074/jbc.M112.369793
20. A.S. Krois, H.J. Dyson, & P.E. Wright, Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain, Proc. Natl. Acad. Sci. U.S.A. 115 (48) E11302-E11310 (2018). https://doi.org/10.1073/pnas.1814051115
21. Yuesheng Zhang, Madhuri Dutta, The multifunctional proline-rich domain of p53 tumor suppressor, Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, Volume 1880, Issue 3, 2025, 189326, ISSN 0304-419X, https://doi.org/10.1016/j.bbcan.2025.189326
22. Capuozzo, M., Santorsola, M., Bocchetti, M., Perri, F., Cascella, M., Granata, V., Celotto, V., Gualillo, O., Cossu, A. M., Nasti, G., Caraglia, M., & Ottaiano, A. (2022). p53: From Fundamental Biology to Clinical Applications in Cancer. Biology, 11(9), 1325. https://doi.org/10.3390/biology11091325
23. Ha J-H, Prela O, Carpizo DR, and Loh SN (2022) p53 and Zinc: A Malleable Relationship. Front. Mol. Biosci. 9:895887. doi: 10.3389/fmolb.2022.895887
24. Oleg Laptenko, David R. Tong, James Manfredi, Carol Prives, The Tail That Wags the Dog: How the Disordered C-Terminal Domain Controls the Transcriptional Activities of the p53 Tumor-Suppressor Protein, Trends in Biochemical Sciences, Volume 41, Issue 12, 2016, Pages 1022-1034, ISSN 0968-0004, https://doi.org/10.1016/j.tibs.2016.08.011
25. Laptenko, O., Shiff, I., Freed-Pastor, W., Zupnick, A., Mattia, M., Freulich, E., Shamir, I., Kadouri, N., Kahan, T., Manfredi, J., Simon, I., & Prives, C. (2015). The p53 C terminus controls site-specific DNA binding and promotes structural changes within the central DNA binding domain. Molecular Cell, 57(6), 1034–1046. https://doi.org/10.1016/j.molcel.2015.02.015
26. Shi, D., & Gu, W. (2012). Dual Roles of MDM2 in the Regulation of p53: Ubiquitination Dependent and Ubiquitination Independent Mechanisms of MDM2 Repression of p53 Activity. Genes & Cancer, 3(3-4), 240–248. https://doi.org/10.1177/1947601912455199
27. Ou, Y. H., Chung, P. H., Sun, T. P., & Shieh, S. Y. (2005). p53 C-terminal phosphorylation by CHK1 and CHK2 participates in the regulation of DNA-damage-induced C-terminal acetylation. Molecular biology of the cell, 16(4), 1684–1695. https://doi.org/10.1091/mbc.e04-08-0689
28. Kim, H.-M., Zheng, X., & Lee, E. (2023). Experimental Insights into the Interplay between Histone Modifiers and p53 in Regulating Gene Expression. International Journal of Molecular Sciences, 24(13), 11032. https://doi.org/10.3390/ijms241311032
29. Qian Hao, Jiaxiang Chen, Hua Lu, Xiang Zhou, The ARTS of p53-dependent mitochondrial apoptosis, Journal of Molecular Cell Biology, Volume 14, Issue 10, October 2022, mjac074, https://doi.org/10.1093/jmcb/mjac074
30. Amundson, S. A., Patterson, A., Do, K. T., & Fornace, Jr., A. J. (2002). A Nucleotide Excision Repair Master-Switch: p53 Regulated Coordinate Introduction of Global Genomic Repair Genes. Cancer Biology & Therapy, 1(2), 145–149. https://doi.org/10.4161/cbt.59
31. Shen J, Wang Q, Mao Y, Gao W, Duan S. Targeting the p53 signaling pathway in cancers: Molecular mechanisms and clinical studies. MedComm. 2023;4:e288. https://doi.org/10.1002/mco2.288
32. Chiang, Y. T., Chien, Y. C., Lin, Y. H., Wu, H. H., Lee, D. F., & Yu, Y. L. (2021). The Function of the Mutant p53-R175H in Cancer. Cancers, 13(16), 4088. https://doi.org/10.3390/cancers13164088
33. Hassin, O., Nataraj, N. B., Shreberk-Shaked, M., Aylon, Y., Yaeger, R., Fontemaggi, G., Mukherjee, S., Maddalena, M., Avioz, A., Iancu, O., Mallel, G., Gershoni, A., Grosheva, I., Feldmesser, E., Ben-Dor, S., Golani, O., Hendel, A., Blandino, G., Kelsen, D., Yarden, Y., … Oren, M. (2022). Different hotspot p53 mutants exert distinct phenotypes and predict outcome of colorectal cancer patients. Nature communications, 13(1), 2800. https://doi.org/10.1038/s41467-022-30481-7
34. Eldar, A., Rozenberg, H., Diskin-Posner, Y., Rohs, R., & Shakked, Z. (2013). Structural studies of p53 inactivation by DNA-contact mutations and its rescue by suppressor mutations via alternative protein-DNA interactions. Nucleic Acids Research, 41(18), 8748–8759. https://doi.org/10.1093/nar/gkt630
35. Zhu, G., Pan, C., Bei, J. X., Li, B., Liang, C., Xu, Y., & Fu, X. (2020). Mutant p53 in Cancer Progression and Targeted Therapies. Frontiers in Oncology, 10, 595187. https://doi.org/10.3389/fonc.2020.595187
36. Olivier, M., Hollstein, M., & Hainaut, P. (2010). TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harbor perspectives in biology, 2(1), a001008. https://doi.org/10.1101/cshperspect.a001008
37. Efe, G., Cunningham, K., Rustgi, A. K., Prives, C., Manfredi, J. J., & Sánchez-Rivera, F. J. (2026). Mutant p53: evolving perspectives. Genes & Development, 40(1-2), 4–25. https://doi.org/10.1101/gad.353408.125
38. Cen Zhang, Juan Liu, Dandan Xu, Tianliang Zhang, Wenwei Hu, Zhaohui Feng, Gain-of-function mutant p53 in cancer progression and therapy, Journal of Molecular Cell Biology, Volume 12, Issue 9, September 2020, Pages 674–687, https://doi.org/10.1093/jmcb/mjaa040
39. Oren, M., & Rotter, V. (2010). Mutant p53 gain-of-function in cancer. Cold Spring Harbor perspectives in biology, 2(2), a001107. https://doi.org/10.1101/cshperspect.a001107
40. Gencel-Augusto, J., & Lozano, G. (2020). p53 tetramerization: at the center of the dominant-negative effect of mutant p53. Genes & Development, 34(17-18), 1128–1146. https://doi.org/10.1101/gad.340976.120
41. Srivastava S, Wang S, Tong YA, Hao ZM, Chang EH. Dominant negative effect of a germ-line mutant p53: a step fostering tumorigenesis. Cancer Res. 1993 Oct 1;53(19):4452-5. PMID: 8402611.
42. A. de Vries, E.R. Flores, B. Miranda, H. Hsieh, C.T.M. van Oostrom, J. Sage, & T. Jacks, Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function, Proc. Natl. Acad. Sci. U.S.A. 99 (5) 2948-2953, https://doi.org/10.1073/pnas.052713099 (2002).
43. Kennedy, M.C., Lowe, S.W. Mutant p53: it's not all one and the same. Cell Death Differ 29, 983–987 (2022). https://doi.org/10.1038/s41418-022-00989-y
44. Wade, M., Li, Y. C., & Wahl, G. M. (2013). MDM2, MDMX and p53 in oncogenesis and cancer therapy. Nature Reviews. Cancer, 13(2), 83–96. https://doi.org/10.1038/nrc3430
45. Li, B., Cheng, Q., Li, Z., & Chen, J. (2010). p53 inactivation by MDM2 and MDMX negative feedback loops in testicular germ

cell tumors. Cell cycle (Georgetown, Tex.), 9(7), 1411–1420. https://doi.org/10.4161/cc.9.7.11255

46. A. Hengstermann, L.K. Linares, A. Ciechanover, N.J. Whitaker, & M. Scheffner, Complete switch from Mdm2 to human papillomavirus E6-mediated degradation of p53 in cervical cancer cells, Proc. Natl. Acad. Sci. U.S.A. 98 (3) 1218-1223, https://doi.org/10.1073/pnas.98.3.1218 (2001).
47. Bernard X, Robinson P, Nominé Y, Masson M, Charbonnier S, *et al.* (2011) Proteasomal Degradation of p53 by Human Papillomavirus E6 Oncoprotein Relies on the Structural Integrity of p53 Core Domain. PLOS ONE 6(10): e25981. https://doi.org/10.1371/journal.pone.0025981
48. Hassin, O., Oren, M. Drugging p53 in cancer: one protein, many targets. Nat Rev Drug Discov 22, 127–144 (2023). https://doi.org/10.1038/s41573-022-00571-8
49. Mandinova, A., & Lee, S. W. (2011). The p53 pathway as a target in cancer therapeutics: obstacles and promise. Science Translational Medicine, 3(64), 64rv1. https://doi.org/10.1126/scitranslmed.3001366
50. Joerger, A. C., Bauer, M. R., Wilcken, R., Baud, M. G. J., Harbrecht, H., Exner, T. E., Boeckler, F. M., Spencer, J., & Fersht, A. R. (2015). Exploiting Transient Protein States for the Design of Small-Molecule Stabilizers of Mutant p53. Structure (London, England: 1993), 23(12), 2246–2255. https://doi.org/10.1016/j.str.2015.10.016
51. Baliakas, P., & Soussi, T. (2025). The TP53 tumor suppressor gene: From molecular biology to clinical investigations. Journal of Internal Medicine, 298(2), 78–96. https://doi.org/10.1111/joim.20106
52. Munisamy, M., Mukherjee, N., Thomas, L., Pham, A. T., Shakeri, A., Zhao, Y., Kolesar, J., Rao, P. P. N., Rangnekar, V. M., & Rao, M. (2021). Therapeutic opportunities in cancer therapy: targeting the p53-MDM2/MDMX interactions. American journal of cancer research, 11(12), 5762–5781. https://pmc.ncbi.nlm.nih.gov/articles/PMC8727821/
53. Wang, S., Zhao, Y., Aguilar, A., Bernard, D., & Yang, C. Y. (2017). Targeting the MDM2-p53 Protein-Protein Interaction for New Cancer Therapy: Progress and Challenges. Cold Spring Harbor perspectives in medicine, 7(5), a026245. https://doi.org/10.1101/cshperspect.a026245
54. Andreeff M, Kelly KR, Yee K, Assouline S, Strair R, Popplewell L, Bowen D, Martinelli G, Drummond MW, Vyas P, Kirschbaum M, Iyer SP, Ruvolo V, González GM, Huang X, Chen G, Graves B, Blotner S, Bridge P, Jukofsky L, Middleton S, Reckner M, Rueger R, Zhi J, Nichols G, Kojima K. Results of the Phase I Trial of RG7112, a Small-Molecule MDM2 Antagonist in Leukemia. Clin Cancer Res. 2016 Feb 15;22(4):868-76. Epub 2015 Oct 12. PMID: 26459177; PMCID: PMC4809642. https://doi.org/10.1158/1078-0432.CCR-15-0481
55. Ray-Coquard I, Blay JY, Italiano A, Le Cesne A, Penel N, Zhi J, Heil F, Rueger R, Graves B, Ding M, Geho D, Middleton SA, Vassilev LT, Nichols GL, Bui BN. Effect of the MDM2 antagonist RG7112 on the p53 pathway in patients with MDM2-amplified, well-differentiated or dedifferentiated liposarcoma: an exploratory proof-of-mechanism study. Lancet Oncol. 2012 Nov;13(11):1133-40. Epub 2012 Oct 17. PMID: 23084521. http://doi.org/10.1016/S1470-2045(12)70474-6
56. Skalniak, L., Kocik, J., Polak, J., Skalniak, A., Rak, M., Wolnicka-Glubisz, A., & Holak, T. A. (2018). Prolonged Idasanutlin (RG7388) Treatment Leads to the Generation of p53-Mutated Cells. Cancers, 10(11), 396. https://doi.org/10.3390/cancers10110396
57. Zhu, H., Gao, H., Ji, Y. *et al.* Targeting p53–MDM2 interaction by small-molecule inhibitors: learning from MDM2 inhibitors in clinical trials. J Hematol Oncol 15, 91 (2022). https://doi.org/10.1186/s13045-022-01314-3
58. Sallman, D. A., DeZern, A. E., Garcia-Manero, G., Steensma, D. P., Roboz, G. J., Sekeres, M. A., Cluzeau, T., Sweet, K. L., McLemore, A., McGraw, K. L., Puskas, J., Zhang, L., Yao, J., Mo, Q., Nardelli, L., Al Ali, N. H., Padron, E., Korbel, G., Attar, E. C., Kantarjian, H. M., … Komrokji, R. S. (2021). Eprenetapopt (APR-246) and Azacitidine in TP53-Mutant Myelodysplastic Syndromes. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 39(14), 1584–1594. https://doi.org/10.1200/JCO.20.02341
59. Thomas Cluzeau *et al.* Eprenetapopt Plus Azacitidine in TP53-Mutated Myelodysplastic Syndromes and Acute Myeloid Leukemia: A Phase II Study by the Groupe Francophone des Myélodysplasies (GFM). J Clin Oncol 39, 1575-1583(2021). https://doi.org/10.1200/JCO.20.02342
60. Cointe, S. M., Korbel, G. A. (2020). Aprea Therapeutics Announces Results of Primary Endpoint from Phase 3 Trial of Eprenetapopt in TP53 Mutant Myelodysplastic Syndromes (MDS). Aprea Therapeutics News Releases. https://ir.aprea.com/news-releases/news-release-details/aprea-therapeutics-announces-results-primary-endpoint-phase-3
61. Mishra A, Tamari R, DeZern AE, Byrne MT, Gooptu M, Chen YB, Deeg HJ, Sallman D, Gallacher P, Wennborg A, Hickman DK, Attar EC, Fernandez HF. Eprenetapopt Plus Azacitidine After Allogeneic Hematopoietic Stem-Cell Transplantation for TP53-Mutant Acute Myeloid Leukemia and Myelodysplastic Syndromes. J Clin Oncol. 2022 Dec 1;40(34):3985-3993. Epub 2022 Jul 11. PMID: 35816664. https://doi.org/10.1200/JCO.22.00181
62. Unknown author. Small molecule restoration of wildtype structure and function of mutant p53 using a novel zincmetallochaperone-based mechanism. Oncotarget. ????; 5: . Retrieved from https://www.oncotarget.com/article/2432/
63. Lindemann A, Patel AA, Silver NL, Tang L, Liu Z, Wang L, Tanaka N, Rao X, Takahashi H, Maduka NK, Zhao M, Chen TC, Liu W, Gao M, Wang J, Frank SJ, Hittelman WN, Mills GB, Myers JN, Osman AA. COTI-2, A Novel Thiosemicarbazone Derivative, Exhibits Antitumor Activity in HNSCC through p53-dependent and -independent Mechanisms. Clin Cancer Res. 2019 Sep 15;25(18):5650-5662. doi: 10.1158/1078-0432.CCR-19-0096. Epub 2019 Jul 15. PMID: 31308060; PMCID: PMC6759991.
64. Shannon N. Westin, Wilberto Nieves-Neira, Christian Lynam, Kowthar Y. Salim, Alison D. Silva, Richard T. Ho, Gordon B. Mills, Robert L. Coleman, Filip Janku, Daniela Matei; Abstract CT033: Safety and early efficacy signals for COTI-2, an orally available small molecule targeting p53, in a phase I trial of recurrent gynecologic cancer. Cancer Res 1 July 2018; 78 (13_Supplement): CT033. https://doi.org/10.1158/1538-7445.AM2018-CT033
65. Azizian NG, Li Y. XPO1-dependent nuclear export as a target for cancer therapy. J Hematol Oncol. 2020 Jun 1;13(1):61. doi: 10.1186/s13045-020-00903-4. PMID: 32487143; PMCID: PMC7268335.
66. Subhash, V.V., Yeo, M.S., Wang, L. *et al.* Anti-tumor efficacy of Selinexor (KPT-330) in gastric cancer is dependent on nuclear accumulation of p53 tumor suppressor. Sci Rep 8, 12248 (2018). https://doi.org/10.1038/s41598-018-30686-1
67. Wang, J., Liu, W., Zhang, L., & Zhang, J. (2023). Targeting mutant p53 stabilization for cancer therapy. Frontiers in Pharmacology, 14, 1215995. https://doi.org/10.3389/fphar.2023.1215995
68. Zhang, Q., Bykov, V. J. N., Wiman, K. G., & Zawacka-Pankau, J. (2018). APR-246 reactivates mutant p53 by targeting cysteines 124 and 277. Cell death & disease, 9(5), 439. https://doi.org/10.1038/s41419-018-0463-7

DOI: 10.36838/IJHSR819.20

69. Roszkowska, K. A., Piecuch, A., Sady, M., Gajewski, Z., & Flis, S. (2022). Gain of Function (GOF) Mutant p53 in Cancer-Current Therapeutic Approaches. International journal of molecular sciences, 23(21), 13287. https://doi.org/10.3390/ijms232113287
70. Shangary, S., & Wang, S. (2008). Targeting the MDM2-p53 interaction for cancer therapy. Clinical Cancer Research: An official Journal of the American Association for Cancer Research, 14(17), 5318–5324. https://doi.org/10.1158/1078-0432.CCR-07-5136

Author

Tanya Verma is a sophomore at Northview High School in Johns Creek, Georgia, United States. She is passionate about disease pathology and aspires to pursue a career in medicine. In her free time, Tanya enjoys swimming and playing the piano.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Supply Without Demand: Japan's Pension Reform and Elderly Male Labor Force Participation

Terry S. Lo

Cardiff Sixth Form College, Trinity Court, 21-27 Newport Rd, Cardiff CF24 0AA, Wales, UK of Great Britain and Northern Ireland; terrlo20090326@gmail.com

ABSTRACT: This paper examines the effect of Japan's 2001 pension eligibility age increase and 2006 Elderly Employment Stabilization Law on male labor force participation for ages 60–64. Using synthetic control and synthetic difference-in-differences with a 27 and 19-country panel dataset spanning 1991–2022, this paper finds no detectable effect of the 2001 pension reform alone, but suggests a positive effect for the 2006 Elderly Employment Stabilization Law. The ATT flips from negative to positive, a swing of nearly 11 percentage points. When using the SDID estimator, the effects of both policies were not significant.

KEYWORDS: Economics, Public Economics, Pension Reform, Japan, Labor Force Participation

Introduction

Japan, with a third of its population over 65, has the highest share of elderly in the world.[1] Moreover, more than 1 in 10 people in the country are now aged 80 or older.[1] This shift in demography is caused by the following three key factors: the post-WWII baby boom between 1947 and 1949, when births exceeded 2.5 million for three consecutive years;[2] a declining fertility rate, which fell to a record low of 1.15 in 2024,[2] well below the replacement rate of 2.1;[3] and a rising life expectancy of their citizens, with Japanese women living an average of 87.13 years, which is the highest in the world and with men living 81.09 years.[4] This rapid aging process piles pressure on Japan's pension system, as a shrinking working-age population has to support a growing number of retirees' income.

Japan's aging population means that there will be fewer workers supporting the pension beneficiaries, as the ratio of working-age people to retirees has experienced a constant decline from 9.5 working-age people per retiree in 1960 to 2.6 by 2009.[5] This placed a growing financial burden on Japan's pay-as-you-go pension system. Therefore, the government enacted pension reform to the eligibility age, which aimed to reduce the total period of benefit payouts, thus mitigating the fiscal pressure on the system.

Japan's pension reform was legislated in 1994 and 2000, increasing the eligibility age from 60 to 65 for flat-rate and wage-proportional pensions, respectively. Both reforms took a gradual approach. The flat-rate component reform was implemented in 2001 for men, with the eligible age increasing one year every three years. However, this created a gap between the new age for the flat-rate component of the pension and the mandatory retirement age, which remained at 60 in most firms, leaving workers without employment and with reduced pension income during those years. To address this, the Elderly Employment Stabilization Law was revised in 2006, requiring firms to retain older workers until the new eligibility age.[6] Unless otherwise specified, 'pension reform' in this paper refers to the 1994 reform that increased the flat-rate component eligibility age, implemented in 2001.[5]

This study hypothesizes that raising the pension eligibility age increased male labor force participation for ages 60–64. The delay of eligibility reduces lifetime pension wealth and makes leisure relatively more expensive, both of which encourage continued work. The theoretical framework section develops this logic further.

Previous studies examined Japan's pension reform using within-country comparisons.[7,8] These studies found that the reform had a positive impact on employment rates, especially after the 2006 law addressed the gap between retirement age and pension eligibility, where it increased employment rates for men aged 60–61 by 2 to 3 percentage points,[7] and that combining the eligibility age increase with demand-side enforcement raised employment at age 63 by approximately 4 percentage points.[7] However, within-country comparisons may be subject to spillover effects. For example, the same eligibility age increase applied to women five years later, meaning that men in supposedly untreated cohorts may have adjusted their labor supply in response to the reduced household pension income from their spouse's delayed eligibility. Cross-country methods avoid this possible contamination. To date, no study has used cross-country comparisons to investigate the effect of the reform. This study addresses this absence of cross-country evidence by using synthetic control and synthetic difference-in-differences to examine the effect of Japan's pension reform on labor force participation. These methods construct a counterfactual using other countries, avoiding within-country contamination.

The pension reform and the 2006 Elderly Employment Stabilization Law affected the labor force participation rate in different ways. The pension reform operates on the supply side by reducing pension wealth and increasing the incentive to work. However, the majority of firms in Japan had their mandatory retirement age set as 60, which created a demand-side constraint, as employers could force workers out regardless of their willingness to continue working.

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.21

To disentangle supply-side from demand-side effects, this paper additionally estimated a 2006 treatment year corresponding to the Revised Elderly Employment Stabilization Law. The key pattern is that the treatment effect flips from negative under the 2001 treatment to positive under the 2006 treatment, a swing of nearly 11 percentage points. This suggests that pension reform required complementary employment legislation to produce labor force participation gains. This finding has broader policy relevance, as many developed countries are currently considering or implementing similar increases to pension eligibility ages.

Institutional Background:

The system:

Japan's public pension system operates on a two-tier structure. The first tier is the National Pension Insurance (NPI, Kokumin Nenkin), which provides a flat-rate basic benefit to all residents regardless of lifetime earnings. The eligibility age for NPI is 65 years old. The benefit from NPI is called the Old-Age Basic Pension. Everyone receives the Old-Age Basic Pension from age 65.

The second tier consists of earnings-related schemes for employees. The Employees' Pension Insurance (EPI, Kosei Nenkin) provides earnings-related benefits to employees; the eligibility age for EPI was 60 years old pre-reform. The Mutual Aid Insurance (MAI) covers government employees and private school teachers. The MAI followed the same eligibility age reform schedule as EPI and was integrated into EPI in October 2015.[5] At 65 years old, 45.5% of the population receive NPI only, 48% receive NPI plus EPI (wage-proportional component), and 6.5% receive the Old-Age Basic Pension plus MAI wage-proportional component.[9] This study focuses on EPI as it covers the largest share of employees. Within the EPI, benefits are divided into two components:

1. Flat-rate component - A fixed amount of money equivalent to the NPI basic pension. This component was designed to substitute for the NPI for retirees under the age of 65. Once they reach 65, the EPI flat-rate component stops and is replaced by the NPI.

2. Wage-proportional component - based on career average earnings. The benefit from EPI is called the Old-Age Employees' Pension. Before age 65, EPI members receive the Special Old-Age Employees' Pension (flat-rate + wage-proportional). From age 65, they receive the Old-Age Basic Pension (from NPI) plus the Old-Age Employees' Pension (wage-proportional component only from EPI).

Before the reforms, EPI beneficiaries could claim both components starting at age 60 through a transitional arrangement called the "Special Payment of the Employees' Old-Age Pension" (TOKUBETSU SHIKYUNO ROREI KOSEI NENKIN).

The reform package (1994 and 2000):

Japan's pension eligibility reform was implemented through two major legislation acts, creating a phased transition that separates the flat-rate component from the wage-proportional component.

1994 Reform Act:

The 1994 Pension Reform Act legislated the increase in eligibility age for the flat-rate component of the EPI from age 60 to 65. The implementation began in April 2001 for male workers, with the eligibility age increasing by one year every three years, as shown in Table 1:[5]

Table 1: Implementation schedule for flat-rate component eligibility age increase (1994 reform, male workers).

Birth Cohort	Eligibility Age	First Affected Year
Born before April 2, 1941	60	-
April 2, 1941 - April 1, 1943	61	2001
April 2, 1943 - April 1, 1945	62	2004
April 2, 1945 - April 1, 1947	63	2007
April 2, 1947 - April 1, 1949	64	2010
Born after April 1, 1949	65	2013

Note: This reform applied to female workers as well, with a 5-year time lag. The flat-rate component eligibility age for females increased with the same pattern starting in 2006.5 The flat-rate eligibility age rises in one-year steps from 60 for cohorts born before April 1941 to 65 for those born after April 1949.

The 2000 Reform Act:

The 2000 Pension Reform Act (passed in March 2000) extended this approach to the wage-proportional component. Implementation began in April 2013 for male workers (and in 2018 for female workers), with the eligibility age increasing by one year every three years from 60 to 65, following the same phased pattern as the flat-rate component (completing in 2025 for men).[5]

The Earnings Test (Zaishoku Rorei Nenkin):

Japan's pension system includes an earnings test (Zaishoku Rorei Nenkin) that reduces pension benefits for workers aged 60–64 whose combined monthly earnings and pension exceed 280,000 yen,[10] with an effective marginal tax rate of 50%.[8] For high earners, pensions are fully suspended. This mechanism interacts with pension reform by reducing the incentive to work while receiving pensions, but it is not the focus of this study.

The 2006 Elderly Employment Stabilization Law:

The Gap Problem. The pension reforms created a growing gap between the flat pension eligibility age and the mandatory retirement age, which remained at 60 in most firms. This gap emerged as a serious social concern,[7] as workers faced a period with neither employment nor sufficient pension income.

The 2006 Legislative Response:

To address this problem, the government revised the Elderly Employment Stabilization Law in 2004, with implementation beginning in April 2006. Beginning with workers who turned 60 in April 2006 (those born in April 1946), employers were legally required to take at least one of the following measures:[6]

1. Raise the mandatory retirement age to the pension eligibility age.
2. Abolish mandatory retirement entirely.
3. Introduce a continued employment system.

Firm Response and Employment Effects. In practice, more than 80% of firms chose the third option,[6] where they offer reemployment contracts to workers after mandatory retirement. These contracts last until the workers reach the pension eligibility age. For example, it is age 63 for those born in 1946, age 65 for those born after April 1949. Reemployed workers usually receive lower wages. Kondo and Shigeoka found that this legal enforcement increased the employment rate for men in their early 60s.[7]

Scenarios. To better understand how the reforms affected different workers, four scenarios are presented below. The experience of each cohort depends on their birth year, pension eligibility age, the gap, and the 2006 law.

The first scenario is NPI only. This applies to workers who are not employed in the private or public sector, such as the self-employed. They are not affected by the EPI eligibility age reforms and receive the Old-Age Basic Pension from NPI starting at age 65. Since they are not subject to mandatory retirement, they do not face the gap problem.

The second scenario is EPI with reform impact but no law addressing the gap. This represents the cohort born between April 2, 1941, and April 1, 1946. Here we use a male born in April 1943 as an example. He turned 60 in 2003; the flat-rate eligibility age was 62, meaning that he could only receive his wage-proportional component. He had to wait until 2005, when he reached 62, to start receiving his flat-rate component. The 2006 law was not yet legislated, so his employer was not required to offer reemployment. This cohort faced the gap: forced out of employment at 60 with reduced pension income for two years.

The third scenario is EPI with reform impact and gap addressed by the 2006 law. This represents the cohort born between April 2, 1946, and April 1, 1949. Here we use a male born in April 1946 as an example. He turned 60 in April 2006 and would have faced mandatory retirement. His flat-rate eligibility age was 63, so at 60, he could only receive the wage-proportional component. The 2006 revision of the Elderly Employment Stabilization Law was now in effect, requiring his employer to offer reemployment until his flat-rate eligibility age, which was 63 in his case. He was therefore rehired at lower wages. At 63, he received both components and ended his reemployment contract. At 65, the flat-rate component was replaced by the Old-Age Basic Pension.

The fourth scenario is EPI with both reforms (flat-rate and wage-proportional delay). This represents the cohort born after April 1, 1949. Here we use a male born on April 15, 1953, as an example. He turned 60 in 2013. The flat-rate eligibility age was 65, and the wage-proportional eligibility age was 61. The 2006 law applied, and he was offered a reemployment contract that lasted until age 65. In 2014, he turned 61 and started receiving the wage-proportional component while continuing to work. In 2018, he turned 65 and retired. He then received the Old-Age Basic Pension (NPI) plus the Old-Age Employees' Pension (wage-proportional from EPI).

Across these four scenarios, the critical determinant of each cohort's experience is the gap between mandatory retirement at 60 and the rising pension eligibility age, and whether the 2006 law was in effect to bridge that gap through employer-mandated reemployment.

Theoretical Framework:

Implicit Tax on Continued Work. Early retirement age (ERA) is the age when Social Security benefits are first available. Social Security Wealth (SSW) is the present discounted value of all future pension benefits a worker is entitled to receive based on their work history. When a worker considers working one more year instead of retiring, their SSW will change. This change is called the accrual.[11]

The accrual can be either positive or negative, depending on the country's pension rules. If a country's pension system offers no reward for delayed claiming of pension, working an additional year will imply a person has to give up one year of pension payments without gaining higher benefits in the future. This creates a negative accrual as SSW is reduced. A negative accrual acts like a tax on working because it reduces the value of an individual's salary, which means part of what they earn will be offset by the pension income that they have given up. For example, a worker at age 60 who would receive 120 annually in pension benefits decides instead to work and earn 360. Of this salary, 120 is effectively offset by the foregone pension, which becomes an implicit tax on continued work. In this example, the implicit tax rate is 120/360, 33% of annual earnings.

When the accrual is expressed as a percentage of annual earnings, it becomes the implicit tax rate. Tax force is the sum of implicit taxes from the early retirement age to age 69.[11] A country with a high tax force penalizes continued work after the early retirement age, which means workers who delay retirement will experience a loss in pension wealth. In contrast, workers in countries with low tax force face little penalty for working longer; therefore, they have more incentive to stay in the workforce.

Substitution and Income Effects. When the minimum age for pension eligibility in Japan increases, two effects operate on labor supply: the substitution effect and the income effect.[12] To illustrate, consider a 60-year-old worker deciding whether to work or retire. Before the reform, if he retired, he received a full pension of 100. If he works, he earns 150 in salary and receives 100 in pension, but the earnings test reduces his pension by 50, giving him 200 total (150 salary + 100 pension – 50 reduction). The gain from working is 100. After the reform, pension eligibility is delayed. If he retires, he receives nothing. If he works, he earns 150 with no pension reduction. The gain from working is 150. The reform increased the return to working from 100 to 150. Leisure (not working) became relatively more expensive (in terms of opportunity cost); therefore, it encourages work. This is the substitution effect.

The income effect operates through lifetime pension wealth. As the eligibility age increases, the total pension wealth a person can receive over their lifetime is reduced, since they receive the same annual amount but starting later. This reduction acts as a negative income shock. Since leisure is a normal good, a decrease in wealth leads individuals to consume less leisure, meaning they work more.

DOI: 10.36838/IJHSR819.21

Japan's Position in the International Context. The NBER International Social Security project examined pension incentives across 12 developed countries.[11] A key finding is the substantial variation in implicit tax rates and how it affects the incentives for working after ERA. As Coile *et al.* state:[11] "Working from the ERA to age 69 results in a loss of SSW equivalent to about 1.6 years of earnings in the US and Japan, the two countries whose systems are the closest to neutral with respect to work. By contrast, in Belgium, France, Italy, and the Netherlands, working from the ERA to age 69 results in a loss of SSW equivalent to 7.2–9.2 years of earnings."[11]

Japan had one of the lowest tax force values among developed countries, with 160%, compared to 800-900% in France, Belgium, and the Netherlands.[11] At 900%, nearly all of a worker's earnings are offset by lost pension wealth, meaning the financial gain from continued work is close to zero. 160% means that the cumulative pension wealth a typical worker forgoes by working from the ERA to age 69 is equivalent to 1.6 years of his earnings. This means Japanese workers faced relatively little penalty for working past the ERA. This low penalty for working longer can be reflected in Japan's relatively high labor force participation rate; approximately 71-76% of men aged 60–64 were in the labor force during the 1990s, compared to less than 20% in France and Belgium.[11]

The Health Capacity Ceiling. Health capacity is one of the limiting factors for the elderly to continue to stay in the workforce. Research on the health capacity to work provides additional context. Wise estimated that approximately 80% of men aged 60–64 have sufficient health to work,[11] by combining the estimated effect of health on employment for younger individuals with the actual health of older individuals in order to predict the capacity to work. As Coile *et al.* report:[11] "the estimated capacity to work (total height of the bar) of men is fairly similar across countries, at roughly 80 percent, while the share actually working (shaded portion) ranges from about 15 to 80 percent."[11] It is worth noting that the countries the quote above is referring to are 12 Western developed countries.

Japan's pre-reform 60-64 cohort's LFPR of 71-76% was relatively high compared to other developed countries, though still several percentage points below this estimated health capacity ceiling.

Implications for Expected Treatment Effect. These considerations suggest that Japan's pension reform may produce a smaller effect than similar reforms elsewhere. With implicit tax rates already low and labor force participation already near the health capacity ceiling, there was limited room for improvement. This does not mean the reform had no effect; rather, it suggests that detecting an effect requires careful consideration of the counterfactual.

Furthermore, supply-side incentives only translate to higher employment if workers can find jobs. Japan's mandatory retirement system was set at age 60 before the 2006 Law, creating a demand-side constraint where even workers who wanted to continue working would be forced to retire. The 2006 Elderly Employment Stabilization Law addressed this by requiring firms to retain workers until the pension eligibility age. This suggests that the pension reform's effect may depend on whether demand-side barriers are addressed alongside.

Many developed countries also reformed their pension systems during this period. If other developed countries were simultaneously reducing their implicit taxes, Japan's LFPR may rise but show no differential effect relative to other reforming nations.

An additional contextual factor is Japan's prolonged economic stagnation during the 1990s. The pre-treatment period (1991-2000) coincides with the Lost Decade, during which GDP growth stagnated, unemployment rose from approximately 2% to 5%,[13] and deflation set in. These macroeconomic conditions likely depressed labor force participation during the pre-treatment period and may affect the counterfactual trajectory estimated by SDID.

Literature review:

Shimizutani and Oshio examined the earnings test for workers aged 65-69 by analyzing the elimination of the earnings test in 1985 and its reinstatement in 2002.[8] They employed the decomposition method proposed by DiNardo, Fortin, and Lemieux[14] with bunching analysis. They found that its elimination in 1985 increased labor supply for the elderly, especially for males, with little evidence that reinstating the test in 2002 affected the labor supply of the elderly. Kondo and Shigeoka focused on workers aged 60–64.[7] They employed a difference-in-differences approach with varying treatment and control cohorts to measure the impacts of the policies. They found that the 2006 Employment Law increased employment rates, and that the effect of raising the pension eligibility age was larger when combined with demand-side enforcement. Okumura and Usui examined how pension reform affected expectations about future benefits,[9] using survey and regression models. They found that individuals responded to the reform by increasing savings as it reduced expectations of future benefits. This suggests that pension reform affects behavior even before workers reach retirement age.

All of these studies used within-country comparisons - comparing treated units with other cohorts in the country. However, within-country quasi-experimental approaches require strong assumptions that policy affecting one group does not spill over to other groups.[11] For example, as the reform took place, workers expecting future pension changes may adjust their employment decisions in advance, potentially contaminating the control group, which might distort the result.

To date, no study has used cross-country comparisons to examine Japan's pension reform. Cross-country methods construct a counterfactual using other countries by giving them weight using data-driven optimization methods, avoiding within-country contamination.

The International Social Security project provides important context for interpreting pension reform effects across countries. They discovered that reforms across developed countries since the 1980s have generally reduced implicit taxes on continued work,[11] which led to rising employment at older ages. It also shows that these reforms explain a substantial share of the increase in older worker employment. Crucially, this was a global

phenomenon among developed nations, not unique to Japan. This suggests that any cross-country analysis of Japan's reform must account for the fact that control countries may also have been reforming their pension systems during the same period.

Methods

To implement synthetic control and synthetic difference-in-differences, a panel dataset is constructed that spans from 1991 to 2022. The panel dataset combines data from various sources. Male labor force participation rates (LFPR) for ages 60–64, the outcome variable, are obtained from ILOSTAT, the International Labor Organization's statistics database.[15] Covariates, which include GDP per capita, unemployment rate, and population aged 65 and above, are drawn from the World Bank World Development Indicators.[13] ILOSTAT provides labor statistics across 189 countries. WDI covers 217 economies.

The original datasets contained over 100 countries. However, many were dropped during data cleaning due to missing values in the outcome variable or covariates, leaving 27 countries with complete data for the study period. The final sample includes 27 countries: Australia, Austria, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, Turkey, and the United States. However, not all countries are suitable to be in the donor pool. Two synthetic control specifications were run, one with 27 and one with 19 countries. Eight countries are excluded from the main specification to ensure the donor pool is comparable to Japan. Chile, Mexico, and Singapore are excluded because their pension systems are structurally different from Japan's pay-as-you-go system: Chile and Mexico operate privatized individual account systems,[16] and Singapore uses a provident fund.[17] Korea is excluded because its pension system was only established in 1988 and did not reach universal coverage until 1999,[18] leaving insufficient pre-treatment data. Turkey is excluded due to high labor market informality. Iceland and Luxembourg are excluded due to small populations (all under 1 million)[13] and distinct labor market characteristics, such as Iceland's near-full employment. Norway is excluded due to its oil wealth. This leaves 19 developed countries as the main donor pool. Unlike Abadie *et al.*, who used SC, which used a subset of pre-treatment outcome years, all pre-treatment outcome years were used as predictors for synthetic control to maximize pre-treatment fit.[19]

This paper focuses on males because the pension reform for females began five years later, leaving less post-treatment data available. The 60–64 age group is selected because this cohort was directly affected by the increase in pension eligibility age from 60 to 65. The treatment year is 2001, when the eligibility age for the flat-rate component first increased from 60 to 61.

Although the reform was implemented gradually over 12 years, 2001 is used as the treatment year because it marks when the policy first began to bind and is the first cohort to face delayed eligibility. Anticipation effects are unlikely to threaten pre-treatment trends because the cohorts aged 60–64 during 1991-2000 were unaffected by the reform. Workers aged 55-59 during 1994-2000 who anticipated the reform may have adjusted their savings or employment plans in advance. However, such anticipatory increases in labor supply would, if anything, decrease post-treatment LFPR, biasing against finding a positive treatment effect. Since the reform is phased in over 12 years, the ATT averages across different treatment intensities, which may underestimate the full reform effect if the effect increases with the size of the eligibility age increase. This approach follows the paper that proposed synthetic control,[19] which similarly treated California's phased Proposition 99 tobacco control program as a single 1988 treatment.

The synthetic control method constructs a counterfactual and compares it with the treated unit.[19] The difference between them is the effect of the reform. The counterfactual is constructed by assigning weights to donor countries to match Japan's pre-treatment LFPR level. Weights are chosen to minimize the difference between Japan and the synthetic control in pre-treatment periods, using both pre-treatment outcome values and covariates (GDP per capita, unemployment rate, and population aged 65+). A combination of countries often provides a better comparison for the region exposed to the intervention than any single country alone. However, when restricted to comparable developed countries, synthetic control fails to achieve adequate pre-treatment fit, where only Switzerland remains in the donor pool with 100% weight; other developed countries, including New Zealand and Denmark, remained in the donor pool but received zero weight. Indicating that Japan is an outlier outside the convex hull of donor countries. Synthetic difference-in-differences is therefore used for further analysis.[20] Unlike synthetic control, SDID uses only the outcome variable and matches on trends rather than levels. The SDID method reallocates weight and attempts to mimic pre-exposure trends using a combination of countries; doing so can weaken the reliance on the parallel trend assumption that traditional DID relies on. The method then assumes that Japan's trend would have been the same as the counterfactual's trend. Standard errors are computed using the placebo variance estimator proposed by Arkhangelsky, Athey, Hirshberg, Imbens, and Wager (2021), which is designed for settings with a single treated unit.

Both methods rely on several assumptions. First, the assumption of no interference between units requires that Japan's treatment does not affect outcomes in donor countries and that no spillover occurs between treated and control units.[19] Second, no similar treatment should be implemented in the donor pool during the study period. For synthetic control, it requires that the treated unit's pre-treatment values can be matched by a weighted combination of donors, which means Japan has to lie within the convex hull of the donor pool. When this condition is violated, which is the case in this study, "we would not recommend using a synthetic control" (p. 495).[19]

Both methods assume no similar treatment occurred in the donor pool. However, many developed countries also reformed their pension systems during the period of study, which affects the interpretation of the estimates.

DOI: 10.36838/IJHSR819.21

Results and Discussion

The increase in eligibility age:

This section presents the results of the synthetic control and synthetic difference-in-differences. Two synthetic controls were run with progressively restricted donor pools. Model 1 includes all 27 countries with complete data. Model 2 excludes Chile, Mexico, Singapore, Korea, Turkey, Iceland, Norway, and Luxembourg, leaving 19 developed countries. Table 2 reports the donor weight for each run, and Table 3 reports the pre-treatment fit.

Table 2: Synthetic control donor weights by model specification.

Country	Model 1 (27)	Model 2 (19)
Mexico	0.429	N/A
Turkey	0.246	N/A
New Zealand	0.089	0
Iceland	0.129	N/A
Chile	0.067	N/A
Norway	0	N/A
Denmark	0.04	0
Switzerland	0	1
All others	0	0

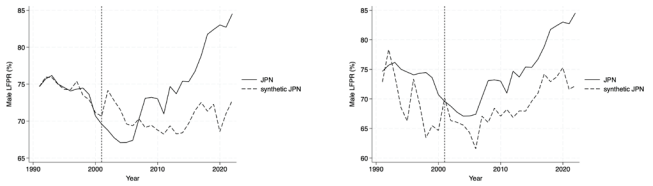
Note: N/A indicates the country was excluded from the donor pool. Weights sum up to 1 within each model. Only countries with non-zero weights in at least one specification are shown. Weight concentrates on emerging-market donors in Model 1 (Mexico, Turkey, and Chile together hold 74.2%), whereas in Model 2, all weight is placed on Switzerland, showing that no combination of developed-country donors can reproduce Japan.

Table 3: Synthetic control pre-treatment fit.

Model	Donor Pool	RMSPE
1	27 countries	0.544
2	19 countries	6.156

Note: RMSPE = Root Mean Squared Prediction Error for pre-treatment period (1991-2000). Lower values indicate a better fit. Pre-treatment RMSPE rises from 0.54 in Model 1 to 6.16 in Model 2, confirming that the restricted developed-country donor pool fits Japan poorly.

Table 2 shows that as the donor pool becomes more restricted, the weights concentrate on fewer countries. In Model 1, weights concentrate on 6 countries: Mexico 42.9%, Turkey 24.6%, Iceland 12.9%, New Zealand 8.9%, Chile 6.7%, and Denmark 4.0%. Notably, emerging markets (Mexico, Turkey, Chile) account for 74.2% of the total weight. In Model 2, the weight is concentrated on one country, Switzerland, with 100%, indicating that Japan cannot be matched by a weighted combination of the remaining developed donor countries. Table 3 further confirms this; it shows the RMSPE increases from 0.54 in Model 1 to 6.16 in Model 2, indicating substantially poorer pre-treatment fit. The good fit in Model 1 (RMSPE = 0.54) is achieved only by relying heavily on emerging markets, which are fundamentally different from Japan in economic structure. While pre-treatment outcome values show excellent fit, structural covariates show substantial imbalance: Japan's GDP per capita is 67% higher than that of synthetic Japan (Model 1), and Japan's elderly population share is more than double that of synthetic Japan. This suggests that Japan lies outside the convex hull of developed country donors; its pre-treatment LFPR level is too distinct to be replicated by any combination of comparable countries. Synthetic control is not recommended when this condition is violated.[19]



(a) Model 1: 27 Countries (b) Model 2: 19 Countries

Figure 1: Synthetic control results by model specification.

Note: The solid line represents Japan. The dashed line represents synthetic Japan. The vertical dashed line indicates the treatment year (2001). Model 1 (27 countries) closely matches Japan before 2001 but relies heavily on emerging-market donors, while Model 2 (19 developed countries) diverges across the whole period, indicating that Japan lies outside of the convex hull of comparable developed countries.

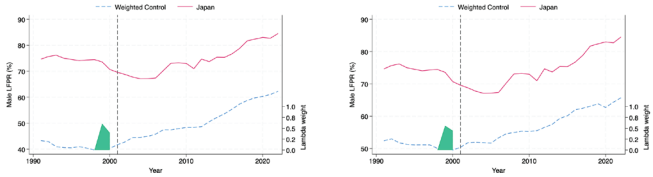
Figure 1 confirms these patterns visually. Model 1 tracks Japan closely pre-treatment, but the post-2001 divergence must be interpreted with caution, given the emerging market composition. Model 2 shows a poor fit, diverging throughout the entire period, confirming that the convex hull condition is violated. Given the convex hull violation, SDID is used as the main estimator.

Table 4: Synthetic difference-in-differences results.

Specification	ATT	SE	p-value	95%CI
19 countries	-8.34	6.65	0.210	[-21.37, 4.70]
27 countries	-5.36	7.79	0.522	[-20.64, 9.91]

Note: ATT = Average Treatment Effect on the treated unit. SE = Standard Error. p-value = probability value for the null hypothesis that ATT = 0. 95% CI = 95% Confidence Interval. The countries include Japan. Both specifications yield negative ATTs (-8.34 and -5.36 percentage points) that are not statistically significant, so no effect of the 2001 reform is detected.

Table 4 presents the SDID results. The ATT for the main specification with 19 developed countries is -8.34 percentage points, with a standard error of 6.65, which indicates that, on average, Japan's LFPR was 8.34 percentage points below what the counterfactual would have predicted after the reform. However, the negative point estimate does not imply the reform reduced labor force participation. The standard error of 6.65 shows that there is uncertainty in the ATT estimate. The p-value of 0.210 is greater than the conventional threshold of 0.05, indicating that the effect is not statistically significant. The 95% CI [-21.37, 4.70] includes zero, implying that the true effect cannot be distinguished from no effect. The wide confidence interval includes both negative and positive values, meaning the true effect could range from a 22 percentage point decrease to a 5 percentage point increase. The robustness check using all 27 countries gives similar results (ATT = -5.36, p = 0.522), confirming that the null finding is robust to a change in the donor pool.



(a) 19 Countries (Main result) (b) 27 Countries (Robustness check)

Figure 2: Synthetic difference-in-differences results.

Note: The red solid line is Japan. The blue dashed line is the weighted control. The green shaded area is the time weights. The vertical dashed line indicates the treatment year (2001). In both donor pools, Japan and the weighted control follow similar upward trends after 2001, with Japan slightly weaker, and the estimated effect is not statistically significant.

Figure 2 depicts the SDID results. In both graphs, Japan and control follow similar upward trends in the post-treatment period, with Japan showing a slightly weaker upward momentum. The two graphs show nearly identical patterns, confirming that the null finding is robust to a change in the donor pool.

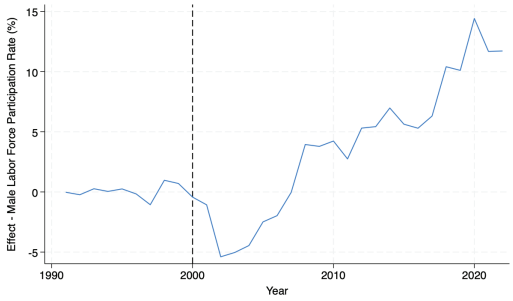


Figure 3: Year-by-year treatment effect (27 countries SC).

Note: Effect is Japan minus synthetic Japan. The vertical dashed line indicates the treatment year (2001). The effect was negative from 2001 to 2006, then turned positive from 2008 onwards, coinciding with the implementation of the 2006 Elderly Employment Stabilization Law.

Year-by-year effects from synthetic control show that from 2001 to 2006, Japan's LFPR was approximately 1.0 to 5.4 percentage points below the synthetic control. Starting in 2008, Japan exceeded synthetic control, reaching approximately 12 percentage points above by 2022. This timing coincides with the 2006 Elderly Employment Stabilization Law, which required firms to retain workers until the pension eligibility age. This suggests that the pension reform alone may have had little effect, but when combined with demand-side enforcement, it produced positive effects. However, individual year effects are not statistically significant ($p > 0.05$ for most years), so this interpretation is suggestive rather than conclusive.

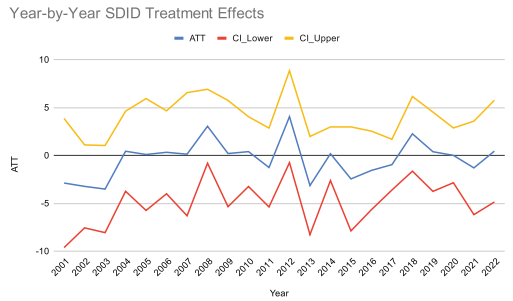


Figure 4: Year-by-year SDID treatment effects (19 countries).

Note: ATT is the year-by-year SDID estimate for the 19-country specification. CI_Lower and CI_Upper are the 95% confidence interval counts. Effects fluctuate around zero with no individual year significant at the 5% level.

Year-by-year SDID estimates show that Japan's LFPR was below the counterfactual in 2001-2003, with effects of -2.88, -3.23, and -3.51 percentage points, respectively, though no individual year was statistically significant at the 5% level. From 2004 onwards, effects fluctuated around zero with no sustained positive or negative trend. The closest to significance was 2012 (ATT = +4.05, p = 0.099). No individual year effect was statistically significant, consistent with the overall null finding.

The year-by-year patterns differ between Figure 3 (synthetic control with 27 countries) and Figure 4 (SDID with 19 countries). In Figure 3, Japan exceeded its synthetic control after 2008, while in Figure 4, effects fluctuate around zero throughout. This discrepancy arises from two sources. Firstly, their donor pool is different; the 27-country SC relies heavily on emerging markets (Mexico 42.9%, Turkey 24.6%), while the 19-country SDID excludes these countries. Secondly, the methods they used are different: SC matches on levels, while SDID matches on trends. The SDID results with the restricted donor pool of developed countries are more appropriate for policy interpretation.

The Elderly Employment Stabilization Law:

To test whether demand-side legislation was necessary for the pension reform to take effect, the analysis is re-estimated using 2006 as the treatment year. This corresponds to the implementation of the Revised Elderly Employment Stabilization Law, which required firms to offer rehire contracts until the pension eligibility age.

Synthetic control with 27 countries assigned weight to Turkey (37.1%), Mexico (26.9%), Iceland (21.9%), Chile (9.4%), and Norway (3.8%), with an RMSPE of 1.09. Figure 5 shows the SC path plot. During the pre-treatment period (1991-2005), there was a negative impact on LFPR due to the Lost Decade. After 2006, Japan's LFPR crossed above the synthetic control and diverged progressively, reaching approximately 18 percentage points above by 2020.

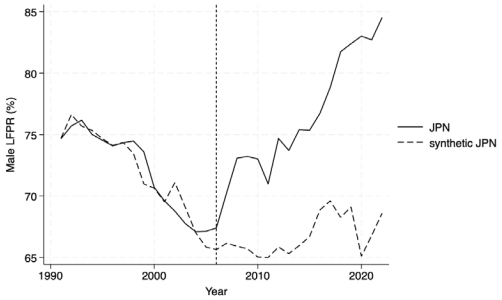


Figure 5: Synthetic control 2006 (27 countries).

Note: The solid line is Japan, and the dashed line is synthetic Japan. The vertical dashed line indicates the treatment year (2006). After 2006, Japan crossed above synthetic control and diverged, reaching approximately 18 percentage points above by 2020.

This sign flip is the most notable pattern in the results. Under the 2001 treatment, the SDID ATT is -8.34 percentage points; under the 2006 treatment, the SDID ATT is +2.50 percentage points (SE = 7.92, p = 0.752). Although the 2006 estimate is not statistically significant on its own, the reversal from negative to positive, a swing of nearly 11 percentage points, is consistent with the interpretation that pension reform alone was insufficient without complementary employment legislation, where pension reform pushed labor supply without addressing demand-side constraints, while the 2006 law provided the employment mechanism through which increased supply-side incentives could translate into actual participation.

DOI: 10.36838/IJHSR819.21

It is important to note that although the sign flip is consistent with the previous result, the size of the coefficients is not significant. This suggests no effect of the policy following the SDID. This leads us to be cautious in interpreting the results from the synthetic control.

Synthetic control placebo-based inference for the 2006 treatment (Table 5) shows effects growing from +4.12 in 2007 to +8.66 percentage points by 2015 (standardized p = 0.208) and +15.93 by 2022 (standardized p = 0.042), which is statistically significant at the 5% level. The gradual emergence of the effect is consistent with the phased implementation of both reforms interacting over time.

Table 5: Placebo Inference for 2006 synthetic control (27 countries).

Year	Effect (pp)	p-value	Standardised p
2006	+1.77	0.708	0.458
2007	+4.12	0.458	0.208
2008	+7.18	0.25	0.125
2009	+7.52	0.208	0.125
2010	+7.98	0.25	0.083
2011	+5.99	0.458	0.167
2012	+8.83	0.375	0.125
2013	+8.39	0.33	0.125
2014	+9.44	0.375	0.167
2015	+8.66	0.417	0.208
2016	+7.86	0.458	0.167
2017	+9.24	0.333	0.208
2018	+13.49	0.167	0.042
2019	+13.28	0.125	0.125
2020	+17.92	0.083	0.000
2021	+15.92	0.083	0.000
2022	+15.93	0.083	0.042

Note: pp refers to percentage points. Standardized p-values divide each unit's post-treatment effect by its pre-treatment RMSPE, following Abadie, Diamond & Hainmueller (2010). The estimated effect grows over the post-treatment period, from +1.77 percentage points in 2006 to +15.93 by 2022, reaching significance on the standardized p-value (p = 0.042) only in the final years.

Robustness check:

Table 6: SDID donor country reform status (male 60–64).

Reform status (during study)	Countries	Combined weight
Major male 60–64 reform	Denmark, France, Germany, Italy, Sweden, Austria, Greece	46.9%
No major male 60–64 reform	Australia, Belgium, Canada, Finland, Ireland, Netherlands, New Zealand, Portugal, Spain, Switzerland, USA	53.1%

Note: Reform status based on whether the country implemented a major reform directly raising the male pension eligibility age for ages 60–64 during the study period. About 47% of donor weight comes from countries that themselves raised the male 60-64 eligibility age during the study period, which motivates the robustness checks that exclude them.

As Table 6 shows, approximately 46.9% of donor weight in the main specification comes from countries that also implemented major reforms raising the male pension eligibility age for ages 60–64 (Austria, Denmark, France, Germany, Greece, Italy, and Sweden), with the remaining 53.1% from non-reforming countries. To ensure the main results are not driven by contamination from these concurrent reforms, the SDID is re-estimated excluding all reforming donors, leaving 12 countries (11 non-reforming donor countries plus Japan). A further specification restricts the donor pool to five countries with no pension reform during the study period (Belgium, Canada, Ireland, New Zealand, and Switzerland), leaving 6 countries, including Japan.

Table 7: SDID Robustness Test.

Specification	2001 ATT	p	2006 ATT	p	Swing
19 countries	-8.34	0.210	+2.50	0.752	+10.84
12 countries	-6.13	0.472	+4.72	0.655	+10.85
6 countries	-4.90	0.322	+3.87	0.454	+8.77

Note: The swing is calculated by subtracting 2001 ATT from 2006 ATT. The countries include Japan. The negative-to-positive sign flip persists across all three specifications, ranging from 8.8 to 10.9 percentage points, suggesting that the pattern is not driven by concurrent reforms in donor countries.

Table 7 reports these specifications. The results across all three specifications are consistent. Each yields negative ATT when the treatment year is set at 2001 and positive ATT when shifted to 2006, with the sign-flip ranging from approximately 8.8 to 10.9 percentage points. This consistency across progressively restricted donor pools indicates that the main findings are not driven by contamination from concurrent reforms in donor countries.

Interpretation and Policy Implications:

Both synthetic control and synthetic difference-in-differences yield no detectable effect on male LFPR for ages 60–64 with treatment year set as the 2001 reform. When the treatment year is moved to 2006, the 27-country synthetic control shows clear divergence between counterfactual and actual Japan - a 15.93 percentage point difference. And as for the SDID, neither estimate is statistically significant on its own. This pattern is consistent with the hypothesis that supply-side pension incentives require complementary demand-side employment policies to translate into gains in the labor force participation rate. This is expected given Section 3's prediction: Japan's low implicit tax force and LFPR already close to the health capacity ceiling meant there was little room for the reform to produce a detectable effect.

As for the underlying reasons behind the null and negative results for 2001, this paper proposes three main factors. First, the low implicit tax force: Japan's tax force was 160%, meaning Japanese workers were already facing very little penalty for working past the Early Retirement Age. Raising the eligibility threshold further reduced the already-small implicit tax; the change in incentive was modest. Second, the ceiling effect: the estimated health capacity ceiling is 80%, and Japan's pre-reform LFPR of 71 - 76% was already close to it. There was comparatively little room for participation to rise compared to other countries, where LFPR among 60–64 year-olds was 15 - 20%. Third, the demand-side constraint. The majority of corporations in Japan had their mandatory retirement age remain at 60 after the 2001 pension reform. This led to a phenomenon where even if workers were incentivized to keep working, employers could and did force them out. Supply-side incentives did not yield expected benefits as there was insufficient demand to absorb the increased supply.

The 2006 Elderly Employment Stabilization Law removed the demand-side constraint by requiring firms to offer re-employment contracts until the pension eligibility age. Workers

who were incentivized by the pension reform to keep working after 60 now had the legal mechanism to do so. The supply-side push could now operate alongside the demand-side pull, and this is supported by Figure 3, which shows the year-by-year SC results. Japan underperformed the synthetic control from 2001 to 2006, then crossed above it beginning in 2008, which is consistent with the expected effect of the 2006 law.

Kondo and Shigeoka reached the same conclusion as this paper using within-country difference-in-differences,[7] showing that the 2006 law increased employment and the pension reform's effect was larger when combined with demand-side enforcement. This paper provides additional suggestive evidence for this finding using cross-country SC and SDID, which are entirely different methods, and yet the same conclusion is reached. The convergence of within-country and cross-country evidence strengthens the credibility of the hypothesis that pension reform requires accompanying employment legislation. Additionally, the NBER International Social Security project noted that most developed countries had their pension systems reformed during the study period, reducing implicit taxes and increasing older worker employment.[11] This suggests that the donor countries in this study were also changing, which helps explain the null result for 2001; Japan's LFPR may have improved, but so did the counterfactual.

Many developed countries are currently raising pension eligibility ages, such as the UK, Australia, and the Netherlands. This finding has implications for other countries: raising the pension age alone may not work if employers can still force workers out at 60. Governments should pair pension reform with employment legislation that requires firms to retain or rehire older workers.

There are several limitations to this paper. Most crucially, neither the 2001 nor the 2006 ATT is statistically significant at the 5% level on the SDID model, which limits the findings to suggestive patterns rather than confirmed effects. The donor pool is comparatively small, which limits statistical power. Among the countries in the donor pool, some also reformed their pension systems during the study period, which may bias the ATT toward zero by contaminating the "no treatment" assumption. Additionally, only approximately half of the 60–64 cohort in Japan were directly affected by the reform, as the eligibility age increase applied to EPI members (private-sector employees), while NPI-only members, such as the self-employed, already faced an eligibility age of 65. The outcome variable, however, captures aggregate male LFPR for the entire 60–64 age group, which may dilute the estimated treatment effect.

Conclusion

This paper used synthetic control and synthetic difference-in-differences to estimate the effect of Japan's 2001 pension reform and 2006 Elderly Employment Stabilization Law on male labor force participation for ages 60–64. The 2001 treatment yielded no statistically significant effect across SC and SDID. However, when the treatment year is moved to 2006, SC shows a 15.93 percentage point increase. This pattern suggests that pension reform alone was insufficient and required the 2006 Elderly Employment Stabilization Law to translate into labor force participation gains. This provides suggestive evidence that countries currently raising pension eligibility age should pair reforms with employer retention mandates or complementary employment legislation.

This paper is the first to use cross-country evaluation on this combination of policy, and it reaches the same conclusion as Kondo & Shigeoka's within-country work. This strengthens the credibility of the finding. For countries that are presently raising or considering raising pension eligibility age to adjust their pension system, the implication is that pension reform without accompanying employer retention mandates risks the same gap issue Japan experienced between 2001 and 2006.

This research is limited by the comparatively small donor pool and by contamination from concurrent pension reforms in donor countries, both of which reduce statistical power. Additionally, only EPI members within the 60–64 cohort were directly affected by the reform, while the outcome variable captures the entire age group, potentially diluting the estimated effect. Moreover, as the reform is phased in, the effect is likely to be diluted over time. Further research using cross-country individual-level data or a larger donor pool could provide more precise estimates of pension reform's effect on elderly employment. A further direction is whether the 2006 law increased elderly participation at the cost of employment quality, as reemployed workers typically accepted substantially lower wages than in their pre-retirement positions. Moreover, the SDID model can yield more accurate results when fed with micro-level data, which presently cannot be obtained from the authorities by me. We should note that because of this, this essay's SDID result should be interpreted with caution.

Acknowledgments

I would like to express my gratitude to Professor Jose Maximo for teaching me how economic theory applies to policy research and for guiding my understanding and utilization of econometrics, bringing economics from paper to real-life application.

References

1. World Economic Forum. *Japan's aging population: The implications for its economy.* https://www.weforum.org/stories/2023/09/elderly-oldest-population-world-japan/ (accessed 2026-02-18).
2. Nippon.com. *Births in Japan Fall Below 700,000 for the First Time in 2024.* https://www.nippon.com/en/japan-data/h02429/ (accessed 2026-02-11).
3. OECD. *Fertility rates (indicator)*. https://www.oecd.org/en/data/indicators/fertility-rates.html (accessed 2026-02-11).
4. Japan Times. Japan's Average Life Expectancy in 2024 Almost Unchanged from 2023, Data Shows. 2025. https://www.japantimes.co.jp/news/2025/07/25/japan/science-health/life-expectancy-average-flat/ (accessed 2026-02-18).
5. Ministry of Health, Labor and Welfare. *The Point of the Pension Plan*; Government of Japan, 2010. https://www.mhlw.go.jp/english/org/policy/dl/p36-37_1.pdf (accessed 2026-02-18).
6. Kondo, A. Effects of Increased Elderly Employment on Other Workers' Employment and the Elderly's Earnings in Japan. *IZA*

DOI: 10.36838/IJHSR819.21

J. Labor Policy **2016**, *5* (2), 1–23. https://doi.org/10.1186/s40173-016-0063-z.
7. Kondo, A.; Shigeoka, H. The Effectiveness of Demand-Side Government Intervention to Promote Elderly Employment: Evidence from Japan. *ILR Rev.* **2017**, *70* (4), 1008–1036. https://doi.org/10.1177/0019793916676490.
8. Shimizutani, S.; Oshio, T. Revisiting the Labor Supply Effect of Social Security Earnings Test: New Evidence from Its Elimination and Reinstatement in Japan. *Jpn. World Econ*. **2013**, *28*, 99–111. https://doi.org/10.1016/j.japwor.2013.08.003.
9. Okumura, T.; Usui, E. Effect of Pension Reform on Pension-Benefit Expectations and Savings Decisions in Japan. *RIETI Discuss. Pap. Ser*. **2011**, No. 11-E-065.
10. U.S. Social Security Administration. *Social Security Programs Throughout the World: Asia and the Pacific, 2018 - Japan*. https://www.ssa.gov/policy/docs/progdesc/ssptw/2018-2019/asia/japan.html (accessed 2026-02-18).
11. Coile, C.; Wise, D.; Börsch-Supan, A.; Gruber, J.; Milligan, K.; Woodbury, R. Social Security and Retirement around the World: Lessons from a Long-Term Collaboration. *J. Pension Econ. Finance* **2025**, *24* (1), 8–30. https://doi.org/10.1017/S1474747223000215.
12. Gruber, J.; Wise, D. A. *Social Security Programs and Retirement around the World: Micro-Estimation*; University of Chicago Press: Chicago, IL, 2004.
13. World Bank. World Development Indicators, 2024. https://databank.worldbank.org/source/world-development-indicators (accessed 2026-02-18).
14. DiNardo, J.; Fortin, N. M.; Lemieux, T. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica* **1996**, *64* (5), 1001–1044. https://doi.org/10.2307/2171954.
15. International Labor Organization. ILOSTAT - International Labour Organization Statistics, 2024. https://ilostat.ilo.org/ (accessed 2026-02-18).
16. Kritzer, B. E. Chile's Next Generation Pension Reform. *Soc. Secur. Bull.* **2008**, *68* (2), 69–84.
17. OECD. *Pensions at a Glance 2007: Public Policies Across OECD Countries*; OECD Publishing: Paris, France, 2007.
18. Korea Development Institute. *National Pension Reform in South Korea*; Korea Development Institute, 2024. https://www.kdi.re.kr/eng/research/focusView?pub_no=18361 (accessed 2026-02-18).
19. Abadie, A.; Diamond, A.; Hainmueller, J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *J. Am. Stat. Assoc*. **2010**, *105* (490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746.
20. Arkhangelsky, D.; Athey, S.; Hirshberg, D. A.; Imbens, G. W.; Wager, S. Synthetic Difference-in-Differences. *Am. Econ. Rev*. **2021**, *111* (12), 4088–4118. https://doi.org/10.1257/aer.20190159.

■ Author

Terry S. H. Lo is presently a Year 12 student at Cardiff Sixth Form College in the United Kingdom. He is an inquisitive student interested in understanding the mechanisms behind economic phenomena and applying economic theory to real-world policy questions. He will continue to navigate the vast landscape of economic academia.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Laminar Flow: A Viable Focus Point for a Sustainable Aviation Future?

Thomas Messeri

St. Louis School Archinto, Via Olmetto 6, Milan, MI, 20123, Italy; thomas.messeri@icloud.com

ABSTRACT: Aircraft designs have always tried to incorporate laminar airflow in order to reduce surface drag, but the recent shift towards sustainability has accentuated the need for reduced emissions, especially in the aviation sector. The European aviation sector alone is expected to deplete its carbon budget compatible with the UN 1.5°C warming goal as early as 2026. Whilst many efficiency gains have been made over the past years, namely the newer composite materials such as those used in the Boeing 787 or the next-generation engines used on Airbus's NEO variants, there is a renewed need for a revolutionary, sector-changing technology, and laminar flow control designs (LFCDs) may provide just this. Primary research in this field focuses on reviewing specific technologies or design elements, such as winglets, and analyzing their effect on aerodynamic efficiency. Very few, however, provide any reference to its maturity or ability to be implemented, where a metric such as the Technology Readiness Level (TRL) can be used. This paper provides an updated review and determination of the feasibility and overall effectiveness of laminar flow control designs to increase aerodynamic efficiency, focusing on commercial aircraft in the short-term future. Although LFCDs provide significant reductions in carbon emissions and, in turn, also decrease contrail production, an extremely harmful by-product of flight that traps heat and is a cause of global warming, they are extremely difficult to implement effectively on a more complex and commercial scale.

KEYWORDS: Engineering, Aerodynamics, Laminar Flow Control, Reynolds' Number, Boundary Layer.

Introduction

Minimizing aerodynamic drag remains one of the most important goals in aircraft design and development, especially with the growing environmental demands on the aviation sector.[1,2]

With pressure being put on the aviation industry to decrease its carbon footprint, the European sector, by the end of 2026, is projected to deplete its allocated carbon budget compatible with the 2050 1.5°C global warming goal,[3] much research has been done in this specific field to find new, innovative ways of decreasing aircraft emissions. For example, modern aircraft have seen incremental improvements in materials and propulsion, yielding moderate efficiency improvements such as those seen in the Boeing 787.[4]

However, a more favorable solution is to decrease the surface drag on the airframe. Recent research shows that every 1% decrease in drag can result in a reduction of 9 million tonnes of CO_2, indicating that small reductions in drag can achieve substantial emissions results.[5] Consequently, laminar flow control has surfaced as a promising research field, presenting potential efficiency improvements that exceed those attainable through gradual advancements in traditional turbulent-flow designs.

A strategy that has been the focus of a large quantity of research for many years is the implementation of laminar flow control designs (LFCD), a technology that looks to reduce skin-friction drag by postponing the shift of fast-flowing air, predominantly over the wings, from laminar to turbulent flow.

Nonetheless, in spite of years of experimental research, the use of LFCDs in commercial aviation is still limited. A large portion of the current literature focuses on small aircraft, such as business jets or specific parts like wings or nacelles. It is also important to note that LFCDs are not the only emerging technology based on drag reduction, and many other technologies have been or are currently being implemented. Namely, winglets or blended wingtips, as seen on newer aircraft like the A350, have been proven to provide around 15% less drag, a huge improvement with a fairly minor change to the actual airframe.[6]

Moreover, recent environmental and aviation research has widened its scope to include more than just fuel use, emphasizing how aircraft-generated contrails contribute to aviation's environmental footprint.[7]

These results point to LFCDs not only enhancing efficiency but also significantly decreasing non-CO2-related climate impacts.

This paper's main objectives are to:

i) State and explain the principles of aerodynamics related to laminar flow that will be subsequently analyzed by focusing on the elements that govern the technology's airliner implementation and scalability potential.

ii) Examine the present state of laminar flow technology within both the engineering field and the commercial aviation sector, highlighting its real-world applications in existing airframes and potential implementations in conceptual aircraft.

iii) Assess the practicality of implementing laminar flow technology in large commercial aircraft in terms of the Technology Readiness Level (TRL) framework developed by NASA.

This evaluation examines both technological advancement and scalability, determining if laminar flow-based designs

© 2026 Terra Science and Education

DOI: 10.36838/IJHSR819.22

can become an extensively used, operational technology. This analysis seeks to provide an assessment of the advantages and drawbacks of laminar flow as a means of enhancing aerodynamic efficiency in commercial aviation.

Most sources reviewed in this paper were found through Google Scholar or digital search engines. Priority was given to peer-reviewed papers and journal articles, whilst technical and experimental reports were used for empirical and quantitative data. These include a 1945 NACA airfoil data report and a 1965 X-21a flight test analysis. The only non-technical source used was an Airbus BLADE program press release, cited only to report test outcomes. Sources were selected only if they specifically applied to past or current subsonic aircraft, and the only literature focused on non-commercial aircraft was the X-21a report, which was included only due to its being a test flight for specific laminar flow control technology.

It is important to emphasize that current literature often labels laminar flow control as either hybrid (HLFC), active (ALFC), or full (FLFC). For clarity, this paper will refer to any non-passive laminar flow control methods, including HLFC, as ALFC, a system that actively stabilizes the boundary layer through suction applied through perforations in specific areas of the airframe.

Section 1 - Aerodynamic Principles

Section 1.1. - Skin Friction Coefficients:

As an aircraft moves through the air, skin friction arises due to viscous shear stress within the boundary layer. To quantify this friction, a laminar skin-friction coefficient is used. For this, an approximated Reynolds' number of 2 × 10^7 for an airliner in a cruise state is used.[8] Using the formula, the average skin-friction coefficient for laminar flow over the whole wing can be calculated as a value of around 3 × 10^-4.

Under the same conditions, a typical turbulent-flow value is around 2.9 x 10^-3 (See Appendix, Equation A1 and A2). These values are, however, derived from flat plate approximations and do not take into account crossflow instabilities or three-dimensional pressure distributions, yet still represent the quantitative incentive for laminar flow over aircraft surfaces.

Overall, the value for turbulent airflow is nearly one order of magnitude greater than that of the laminar flow, meaning that maintaining even small amounts of laminar airflow over a section of a wing is disproportionately beneficial in reducing total aircraft surface drag.

Section 1.2. - Reynolds' Number:

Reynolds' number is a fundamental tool used in the field of aeronautical engineering, and it represents the ratio between inertial and viscous forces in the flow of a fluid, helping to determine whether airflow, in this case, is turbulent or laminar.

Reynolds' number can be defined with the formula:

$$Re = \frac{\rho V x}{\mu} \tag{1}$$

Where ρ is the density of air, V is the velocity, x is the length at which transition occurs, and μ is the dynamic viscosity of air.

This formula can be used to determine the location of transition along the chord (the distance between the leading and trailing edges). Using a transitional Reynolds' number value of around 5 × 10^6 and substituting with typical aircraft cruise values:[8] ρ = 0.385 kg m^{-3}, V = 230 m s^{-1}, μ = 1.45×10^{-5} kg m^{-1}s^{-1}.

Rearranging and solving for x gives a value of $x \approx 0.82$m.

Given that an Airbus A320's chord length is approximately 4m,[9] the transition to turbulent flow would take place very early. This value results in around 20% laminar flow, highlighting the immediacy with which airflow loses its laminar properties on normal airliners.

Suction-based ALFC aims to achieve laminar flow over around 45% of the chord,[10] which, applied to the context of the A320, is around 1.8m along the chord. Substituting this value into Equation 3 gives a Re of approximately 1.1 × 10^7, more than double the initial boundary layer stabilization.

Section 1.3. - Boundary Layer Thickness:

Using the previous calculations, the boundary layer thickness at the transition point can be approximated using the formula:

$$\delta_{lam} = \frac{4.91x}{\sqrt{Re}} \tag{2}$$

Substituting gives a value of around 0.002 m (2 mm), indicating just how prone the boundary layer is to any minimal contamination or surface roughness.

Section 1.4. - Pressure Gradients:

Another factor that affects the stability of the boundary layer is the pressure gradient along the wing.[11] Starting with the pressure coefficient formula:

$$Cp = \frac{p - p_\infty}{\frac{1}{2}\rho V^2} \tag{3}$$

Where p is pressure, p∞ is the pressure in the freestream, ρ is the air density, and V is the air velocity. Rearranging for ρ and differentiating to find the pressure gradient, seeing as p∞ is constant, then gives:

$$\frac{\mathrm{d}\rho}{\mathrm{d}x} = \frac{1}{2}\rho V^2 \times \frac{\mathrm{dCp}}{\mathrm{d}x} \tag{4}$$

To illustrate this, using an author's own estimated derivative of Cp of -0.5/m from data about a 65-series airfoil, documented graphically in a NACA report analyzing various airfoil characteristics,[12] and using the constant values for air density and velocity, the pressure gradient is roughly -5000 Pa m^{-1} (See Appendix, Equation A3). This negative pressure gradient is favorable, which accelerates the airflow and stabilizes the boundary layer. This illustrative figure of -5000 Pa m^{-1} remains consistent with experimental tests such as the BLADE demonstrator, further analyzed in Section 2.2, and with the sustained laminar flow over 45% of the chord figure discussed in Section 1.2.[10]

Laminar flow airfoils are therefore designed with a favorable pressure gradient in mind, meant to sustain it over much of the forward chord. Shifting the hybrid suction systems towards the aft section of the wing, near the trailing edge, would, in theory, suppress instability growth by reducing the inflection

in the velocity profile.[13] However, a design such as this one would inherently be extremely sensitive to different types of surface contamination and imperfections, having an unintended effect on the pressure gradient, potentially causing a transition to turbulent flow, and rendering its use impractical.

Section 1.5. – Summary:

The equations and calculations presented in this section highlight numerous implications and dependencies for ALFC implementation:

i) There is a large difference between surface drag produced with laminar versus turbulent airflow, explaining how even minimal ALFC application can be extremely beneficial.

ii) The underlying importance of operational service smoothness over surfaces that produce laminar flow, mostly wings, is due to the minute boundary layer thickness that is created.

iii) Laminar flow is achieved and maintained not only through airflow manipulation but also through the creation of pressure gradients that stabilize the boundary layer and contribute to a delayed transition.

This quantitative information provides a basis for a feasibility assessment of the LFCD technologies, explored in a further section.

■ Section 2 - Design Strategies

Section 2.1. – Laminar Flow Control Technology:

The main goal of laminar flow control devices as aerodynamic design strategies is to reduce the skin-friction drag on certain areas of the airframe, mostly the airfoils such as the wings or horizontal stabilizers, by either delaying or, ideally, suppressing the transition from laminar to turbulent air, as shown in Figure 1 below. Seeing as the air flows in layers over the surface, the objective is to keep the boundary layer stable, the layer of air that is in contact with the aircraft, to stop the flow from becoming turbulent. This is done because the skin friction drag due to this turbulent air is the main cause of air resistance in an airliner's cruise state. Even very small improvements in delaying the transition can proportionally have much larger effects on drag reduction due to the airliner's greater size, as examined in terms of Reynolds' number in Section 1.2.

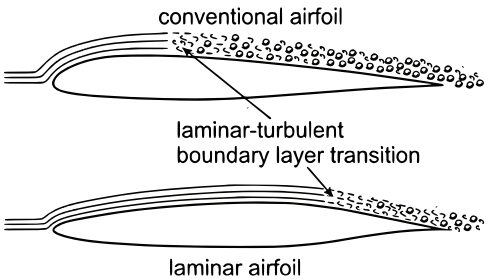


Figure 1: Laminar to turbulent transition on a laminar airfoil compared to a conventional airfoil - adapted from "Novel aeroelastic instabilities on laminar wings" by DLR.[14]

The term 'laminar flow control' encompasses a wide set of different strategies. These are most often categorized into 3 different approaches: Natural Laminar Flow (NLF), Active Laminar Flow Control (ALFC), and Full Laminar Flow Control (FLFC). They vary in both complexity and effectiveness. NLF is on the lower end of both of these scales, due to its nature as a passive approach, whilst active, hybrid, or fully controlled designs use different mechanical systems to stabilize the boundary layer.

Section 2.2. – Natural Laminar Flow:

Natural Laminar Flow (NLF) is the most widely implemented form of LFCD, relying on passive aerodynamic design instead of active boundary layer stabilization. The careful shaping of airfoils helps maintain a favorable pressure gradient, where pressure decreases in the direction of the flowing air, allowing the air to accelerate and the flow to remain attached to the surface. This delays the transition to turbulent air, prolonging laminar flow without the need for additional mechanical systems.

The most common way to increase NLF is to keep a seamless aerodynamic profile across the whole airframe, so as not to disrupt the flow of air. This can be done, for example, by integrating various structural components such as rivets or seams, requiring high-precision manufacturing. Another key element is the wing design, as NLF-based aircraft often make use of extended leading edges. This is because the extended section of the rounded surface divides the airflow softly, minimizing disruption to the boundary layer. The rounded leading edge also allows the stagnation point (where the airflow splits) to move as the angle of attack changes, helping it maintain attached flow in a variety of flight conditions, a necessity for commercial aircraft. It should be noted, however, that this is only applicable to subsonic aircraft since rounded leading edges are likely to cause shock waves at supersonic speeds.

Currently, NLF has been most effective when incorporated into airframes that function at relatively low Reynolds' numbers, where the boundary layer is naturally stable. Consequently, the focus on NLF is extensive in gliders, due to their heavy reliance on aerodynamics, and on several general aviation aircraft, in order to minimize fuel burn. In these situations, which have lower airspeeds, less turbulent atmospheric conditions, and smaller surfaces, NLF has been shown to deliver significant reductions in surface drag, translating to improved range.

However, NLF is much easier to achieve and maintain on these smaller scales. In the commercial aviation sector, whilst inspiration is taken from NLF when designing certain aerodynamic components (wings, stabilizers, tail, etc.), NLF has been limited to restricted surface areas, and these implementations usually only attain partial laminar flow. With larger aircraft and faster cruise speeds, Reynolds' numbers increase significantly, making the boundary layer much less stable and therefore more difficult to sustain constant laminar flow. NLF's huge susceptibility to atmospheric and weather conditions is another weakness. Even slight surface contamination, such as ice accumulation or manufacturing imperfections, can initiate the transition to turbulent air, hindering NLF's overall effectiveness. All of these factors restrict NLF's use to specific areas of the airframe.

DOI: 10.36838/IJHSR819.22

The most influential and groundbreaking test was the Airbus BLADE demonstrator program, where NLF was tested using a modified wingtip section, integrated into an existing airliner, the A340-300. At the termination of the project, Airbus reported that the series of tests was highly successful, with a 50% reduction in wing-related skin-friction, leading to around 5% lower CO_2 emissions.[15]

Despite these constraints, NLF is still vital to the design and development of aircraft, offering a more simplistic approach to aerodynamic efficiency and serving as a base upon which sophisticated laminar flow control methods can be developed.

Section 2.3. - Active Laminar Flow Control:

The inherent limitations of NLF at higher Reynolds numbers motivate the development of active laminar flow control (ALFC) systems. This strategy serves as a middle ground between passive laminar flow systems and a theoretical full laminar flow control system. As previously mentioned in Section 2.2., ALFC builds upon NLF designs that enhance favorable pressure gradients, with active boundary-layer suction systems implemented through small perforations in the airframe. These suction holes extract low-energy air from the boundary layer, inhibiting the creation of disturbances that cause turbulence, allowing laminar flow to be sustained over greater areas. This area can be seen in Figure 2 by the section labeled X_{cr}.

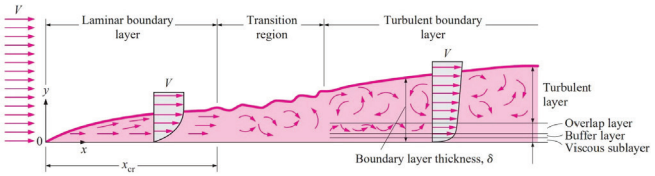


Figure 2: Illustration of the boundary layer during the laminar to turbulent transition, adapted from Fluid Mechanics: Fundamentals and Applications by Y. Cengel.[11]

ALFC's main benefit is the capacity to maintain laminar flow under operating conditions where NLF fails, mostly seen at higher Reynolds numbers common in airliners. ALFC is appealing to airliners since NLF methods prove to be insufficient, and a full laminar flow control system would be too complex to integrate. Additionally, the ability for ALFC to be incorporated only in certain critical areas gives it an almost modular nature, meaning it can adapt to various airframe designs.

In practice, ALFC is still in its experimental phase. Numerous flight tests have been made, each mostly concentrating on one individual component, such as the wing leading edges or the engine nacelles. The latter has proven to be very effective due to its axial symmetry, leading to smoother pressure gradients that support laminar flow well.

ALFC, however, still presents considerable engineering difficulties due to its structural weight and complexity, which could nullify the drag reductions gained. In addition, just like with NLF, weather-related contamination, such as ice accumulation, can impair the ALFC's performance. These issues are the ceiling to what this technology can achieve in individual areas; a ceiling that can be removed by applying ALFC across greater aircraft sections to increase aerodynamic effect, a technology known as FLFC.

Section 2.4. - Full Laminar Flow Control:

Full laminar flow control (FLFC) is the most ambitious and advanced method of boundary-layer management, by sustaining laminar flow over extensive areas of an aircraft, such as an entire wing section. By applying suction on the whole surface, FLFC could hugely decrease the chance of airflow disturbance and greatly delay the turbulent air transition, especially at high Reynolds numbers.

The closest example there has been to an FLFC-based aircraft was the X-21a, an experimental airplane developed by Northrop and used by NASA from 1963 to 1968. It was a modified Douglas B66 aircraft, with replacement engines, reduced wing sweep, and around 800,000 slots through which boundary-layer airflow was sucked, resulting in laminar flow over the wings.[16] Whilst it brought in new data that could not otherwise be obtained, the project was eventually canceled due to its sensitivity to conditions and contamination.

Given this, FLFC is yet to take concrete form. Testing so far has consisted of wind tunnel experiments or digital simulations.[17] These investigations have shown that continuous laminar flow can be attained under precisely regulated conditions, confirming what was speculated. Nonetheless, practical implementations have been limited to small-scale models or standalone test surfaces instead of full airframe configurations.

The main challenge that FLFC faces is its own technical demands, making it difficult to design and integrate into an aircraft. Ongoing suction requires elaborate networks of holes, internal piping, and pumping systems, all of which elevate structural weight and energy usage. This is why development so far has been confined to separate elements that have not been incorporated into real-world aircraft.

Additionally, just like for the other laminar flow strategies, the performance of FLFC is most responsive to surface conditions, and even slight surface imperfections can greatly interfere with the smooth airflow.

Due to these limitations, FLFC is currently seen as more of a theoretical benchmark used to evaluate other laminar flow approaches, rather than a short-term solution for commercial aviation. The actual future integration of FLFC will be later assessed in detail in terms of the TRL scale.

Section 2.5. - Application by Aircraft Component:

Laminar flow control technologies are concentrated on particular aircraft components, those where laminar flow is most advantageous and easiest to sustain.

Specifically, the upper-wing surface during cruise conditions presents considerable opportunities for reducing skin-friction drag if laminar flow can be maintained over a large enough chord distance. The leading edge tends to be the preferred wing section for NLF and ALFC concepts, where beneficial pressure gradients can be created. Nonetheless, wings are still one of the most difficult elements to engineer around laminar flow management due to the extreme aerodynamic forces and

their large area, making it a surface that undergoes major contact with precipitation and environmental contamination.

Another promising option for LFCD placement is on the engine nacelles, especially for hybrid methods. Their smooth and axisymmetric shape results in very stable flow conditions compared to other surfaces, such as wings, and the pressure gradients are often even more favorable.[18] With this in mind, nacelle NLF and ALFC have been a focal point for many testing programs and experimental initiatives, such as those undertaken by Hurel-Dubois and Rolls-Royce.[19] This makes nacelles an appealing solution for the initial implementation of laminar flow technologies in commercial aviation.

The empennage elements, such as horizontal and vertical stabilizers, provide several advantages when compared to other surfaces. These components function, due to their size, at much lower lift coefficients and have decreased pressure gradients, making them a more stable environment for the boundary layer. Additionally, their smaller dimensions and aft position on the airframe create a decreased risk of contamination from external sources. Whilst this makes them ideal, their effectiveness is the least significant of the three components that this paper presents, not only due to the reduced airflow that the surfaces receive, but also the complications of implementing flow control systems within a much smaller volume.

Section 3 - Discussion and Feasibility Study

Section 3.1. - The TRL Framework:

Whilst laminar flow control designs all promise drag reduction and fuel savings when compared to conventional wings and other aerodynamic strategies, their successful implementation on full-scale airliners relies on finding the correct balance between aerodynamic efficiency, operational reliability, and financial feasibility. This can be evaluated using the Technology Readiness Level (TRL) framework created by NASA during the 1970s. The TRL structure is a systematic method to assess and evaluate the maturity of various technologies, broadly utilized in the aerospace research and development sector. The TRL is based on a scale from 1 to 9, where TRL 1 is a mere observation of fundamental physics principles, and TRL 9 is a validated technology in its final form under full use in an operational environment.[20] This framework relates especially well to laminar flow control designs as it takes into account all aspects, not just aerodynamic effectiveness. This also means that each of the various strategies discussed in Section 1, whilst they are based on the same physics foundations, has a different feasibility level due to complexity and integration difficulties. It must be noted that, even if the TRL structure does give an overall view of technology readiness, it falls short in highlighting certain elements, notably the varying cost, a factor that affects the ability to implement the technology.

Section 3.2. - TRL Assessment of NLF:

Currently, Natural Laminar Flow (NLF) is the most advanced laminar flow design. In some cases, it could be given a TRL level of 9, due to its present incorporation into operating aircraft. As analyzed, NLF depends on passive aerodynamic design techniques, making it straightforward to implement. This should make airworthiness certification and regular maintenance simple to achieve. NLF has been implemented on many general aviation aircraft, such as the HondaJet, and has been used on several components, such as engine nacelles, of mainstream airliners such as the Boeing 737 Max and 787, where Reynolds' numbers remain low enough to sustain pressure gradients suitable for laminar flow.[21] However, whilst it has been fairly successful in low Reynolds' number environments and on small components, it is yet to be deployed on extensive surfaces of large aircraft since, as demonstrated in Section 1.2, airliner chord Reynolds' numbers exceed 2 × 10^7 leading to laminar-turbulent transition within the first 20% of the chord. Here, passive NLF shaping is insufficient to maintain boundary-layer stability that would constitute meaningful drag reduction. This element of NLF should remain at a TRL level 7, having been demonstrated in an operational environment, such as with Airbus BLADE. Despite the aerodynamic challenges faced and its lack of validation at a production scale for commercial airliners, NLF has proven to be resilient in real-world scenarios and has brought moderate improvements in skin-friction drag reduction.

Section 3.3. - TRL Assessment of ALFC:

Active Laminar Flow Control (ALFC) is situated, according to this analysis, at an intermediate stage of the TRL framework, at TRL 5-6, depending on the individual system. ALFC integrates passive aerodynamic shaping with active suction and has been the subject of many tests both in wind tunnels and in controlled flight conditions, confirming its aerodynamic advantages at relevant Reynolds numbers. This means it meets TRL standards concerning functional performance. Additionally, programs like the Airbus BLADE demonstrator have meant that ALFC has proven beneficial in an operational environment.[22] However, it does not meet TRL 7 requirements due to the system not yet being scalable, with issues throughout the supply chain. Initially, at the manufacturing level, the high system complexity means the production of composite material components is intensive, considering surfaces must be built and maintained at a near-perfect smoothness. As established in Section 1.3, the boundary layer is approximately 2mm thick, meaning that even irregularities at a millimeter scale are enough to trigger turbulent transitions. This requisite precision makes long-term reliability and maintenance compatibility within the airline sector both complex and expensive, especially considering maintenance would need to be more thorough and frequent, causing possible scheduling problems across airline fleets. Due to such reasons, commercial manufacturers, along with airlines, are yet to commit to the implementation of such technology.[23] All of these problems have an underlying economic obstacle that hinders advancement past this TRL stage.

Section 3.4. - TRL Assessment of FLFC:

Full Laminar Flow Control (FLFC), designed to sustain laminar flow over most or all of a surface through continuous suction, is assigned by this paper a relatively low TRL, around

DOI: 10.36838/IJHSR819.22

TRL 2–3. Although the fundamental principles behind FLFC are clearly understood and have been confirmed through both theoretical studies and experimental tests, this strategy has only been conducted in demonstrations in a laboratory setting. The key link that is missing in order for this technology to climb up the TRL scale is the lack of application on real-world aircraft. For now, the low TRL is due to outstanding issues concerning suction power needs and the structural integration of such technology. In contrast to ALFC, where active control is restricted to specific areas, FLFC requires complete functionality for the entire duration of operation to achieve notable gains in skin-friction. As a result, even with its immense potential for drag reduction, FLFC is still far from achieving the benchmarks needed for a higher TRL.

Section 3.5. - Future Pathways:

Looking to the future, advancements will certainly be made that permit the implementation of laminar flow control. For Active Laminar Flow Control, advancing into the higher end TRLs relies on the technological ability for flight testing that confirms operational durability in a working environment. Global policy and certification changes will also be necessary for ALFC technology to be utilized. However, likely, Full Laminar Flow Control will not progress as much in the short term. This is due to easier development routes that bring more immediate results in reducing drag and fuel consumption, such as winglets or surface coating, down to their lower system complexity. Consequently, it is expected that FLFC will remain in a research phase in restricted conditions until the difficulties associated with its integration can be solved. Significant certification issues also lie in the way of ALFC and FLFC implementation, with policy changes required in order to allow airworthiness. Most importantly, manufacturing and maintenance quality standards would need to be introduced, focused on maintaining minimal surface irregularities, as no current EASA or FAA standards exist for laminar flow surface quality. This would ensure operational reliability, something that would have to be taken into account when considering fuel planning and contingency regulations. Current reserve fuel percentages are around 3-5%, inadequate to cover a potential full loss of laminar flow over an extended cruise period.[24]

Addressing TRL's economic limitation, according to a NASA computational study, a laminar flow wing design could achieve fuel savings of up to 10% on long-range commercial aircraft annually, equating to savings of millions of dollars per aircraft.[25] The optimistic introduction of such FLFC design would likely be similar to the increased costs of manufacturing and maintenance, meaning profitability would depend on how well controlled the operations of the industry are and how scalable the actual technology is. ALFC designs are unlikely to provide such large savings whilst they would still create similar increased costs, meaning, at least during the short term, it is improbable that their deployment would be financially advantageous.

■ Conclusion

This paper set out three main objectives: to explain the aerodynamic principles that govern laminar flow technology and its implementation; to examine the current state of LFCDs in commercial aviation; and to assess their readiness and scalability for the future using the NASA TRL framework. Overall, these objectives reveal that, whilst laminar flow control is well proven and understood, it is not yet industrially mature enough for commercial aviation.

Regarding the technology's present state, after examining the technological and aerodynamic principles behind different laminar flow control methods and the feasibility of their implementation in commercial aviation, this analysis confirms that there are several meaningful advantages in terms of drag reduction and associated fuel burn, a vital solution to reducing aviation's environmental footprint. Whilst these benefits have several practical limitations due to surface and condition sensitivity, when utilized effectively, they provide disproportionate gains in aerodynamic efficiency.

Using the Technology Readiness Level metric, different laminar flow control strategies have major distinctions. Natural laminar flow is operationally mature and in service on several aircraft, albeit at a component level on commercial aircraft. Active laminar flow control has shown promising performance in full-scale testing, representing the most realistic short-term pathway, yet its current characteristics sustain systematic, rather than aerodynamic, barriers involving maintenance and certification. Full laminar flow control, although it provides the most substantial theoretical benefits in airflow management, remains at a low TRL, with issues regarding structural implementation. There would also be the need for an introduction of new airworthiness standards that take into account new redundancies and failure behavior that such technology would create. Maintenance standards would also have to be updated, with strict operational procedures that prioritize safety. Only then will it be plausible for ALFC and FLFC systems to be used in commercial airliners.

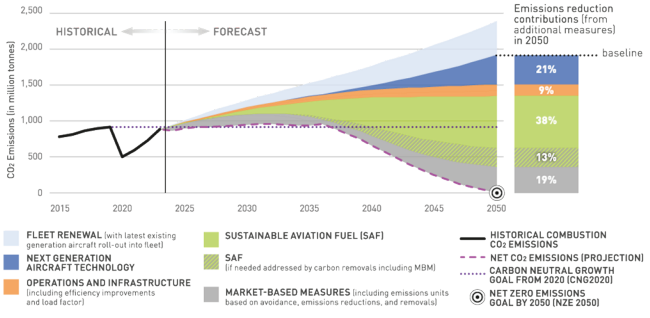


Figure 3: Future projection of aviation emission reductions by method. Adapted from Waypoint 2050 by ATAG.[26]

Future progress within this field of technology is still likely to be gradual, as the many systems involved are made more effective, enabling progressive implementation into existing aircraft. According to the third edition of Waypoint 2050, as seen in Figure 3, without the introduction of any sustainability-focused technology, the commercial aviation sector would

generate approximately 2200 million tonnes of CO_2 emissions.26 Also in Figure 3, next-generation aircraft technology, including laminar flow control, is expected to reduce emissions by 21% by 2050. Drawing from both of these values shows that laminar flow control devices could contribute, alongside the introduction of other technologies, to a total saving of 460 million tonnes of CO_2 emissions per year. In the context of modern passenger jets, laminar flow control therefore provides an appealing and credible pathway towards contributing to increased efficiency and a reduction of commercial aviation's global environmental impact.

Acknowledgments

I would like to acknowledge and thank my IRIS professor, TA, and my parents for their support and guidance throughout the writing of this paper.

References

1. Fleming, G. G.; Marthe, C.; Foca, S.; Pace, D. Environmental Trends in Aviation to 2070. **2025**.
2. Gonzalez-Garay, A.; Heuberger-Austin, C.; Fu, X.; Klokkenburg, M.; Zhang, D.; Made, A. van der; Shah, N. Unraveling the Potential of Sustainable Aviation Fuels to Decarbonize the Aviation Sector. *Energy Environ. Sci.* **2022**, *15* (8), 3291–3309. https://doi.org/10.1039/D1EE03437E.
3. Transport & Environment. Aviation Industry Plans for Growth Irreconcilable with Europe's Climate Goals. **2025** https://www.transportenvironment.org/articles/aviation-industry-plans-for-growth-irreconcilable-with-europes-climate-goals (accessed 2026-05-13)
4. Officer, C. T. *The Boeing 787 Dreamliner: Designing an Aircraft for the Future*. Journal of Young Investigators. **2010** https://www.jyi.org/2010-august/2010/8/6/the-boeing-787-dreamliner-designing-an-aircraft-for-the-future (accessed 2026-01-15).
5. Radespiel, R. *Drag Reduction by Laminar Flow Control. Energies*. **2018** https://doi.org/10.3390/EN11010252.
6. Gavrilović, N. N.; Rašuo, B. P.; Dulikravich, G. S.; Parezanović, V. B. Commercial Aircraft Performance Improvement Using Winglets. *FME Trans.* **2015**, *43* (1), 1–8. https://doi.org/10.5937/fmet1501001G.
7. Teoh, R.; Schumann, U.; Gryspeerdt, E.; Shapiro, M.; Molloy, J.; Koudis, G.; Voigt, C.; Stettler, M. E. J. Aviation Contrail Climate Effects in the North Atlantic from 2016 to 2021. *Atmospheric Chem. Phys.* **2022**, *22* (16), 10919–10935. https://doi.org/10.5194/acp-22-10919-2022.
8. Redeker, G.; Horstmann, K. H.; Koster, H.; Quast, A. Investigations on High Reynolds Number Laminar Flow Airfoils. *J. Aircr.* **1988**, *25* (7), 583–590. https://doi.org/10.2514/3.45627.
9. Airbus S.A.S. *A320 Aircraft Characteristics: Airport and Maintenance Planning*, Rev. 44; Airbus S.A.S.: Blagnac, France, **2024**. https://www.aircraft.airbus.com/sites/g/files/jlcbta126/files/2025-01/AC_A320_0624.pdf (accessed May 17, 2026).
10. Parikh, P. G.; Lund, D. W.; George-Falvy, D.; Nagel, A. L. Hybrid Laminar Flow Control Tests in the Boeing Research Wind Tunnel; **1990**. https://doi.org/10.4271/901978 (accessed 2026-01-23)
11. Çengel, Y. A.; Cimbala, J. M. *Fluid Mechanics: Fundamentals and Applications*; McGraw-Hill series in mechanical engineering; McGraw-Hill Higher Education: Boston, Mass., **2006**.
12. Abbott, I H.; Von Doenhoff, A E.; Stivers Jr., L S. Summary of Airfoil Data Report No. 824. National Advisory Committee for Aeronautics. **1945** https://ntrs.nasa.gov/api/citations/19930090976/downloads/19930090976.pdf (accessed 2026-02-03)
13. Chen H.C.; Patel V. C. *Laminar flow at the trailing edge of a flat plate | AIAA Journal*. **1965**. https://doi.org/10.2514/3.9723 (accessed 2026-01-20).
14. Braune, M. *Novel aeroelastic instabilities on laminar wings. DLR Technical Article* **2021** https:// www.dlr.de/en/ae/latest/technical-articles/technical-articles-2021/novel-aeroelastic-instabilities-on-laminar-wings (accessed 2026-01-27).
15. *Airbus' BLADE demonstrator program wins the Aviation Week Laureates Award for Technology | Airbus*. **2018** https://www.airbus.com/en/newsroom/press-releases/2018-03-airbus-blade-demonstrator-programme-wins-the-aviation-week (accessed 2026-02-01).
16. Kosin, R. E. Laminar Flow Control by Suction as Applied to the X-21a Airplane. *J. Aircr.* **1965**, *2* (5), 384–390. https://doi.org/10.2514/3.43672.
17. Scholz, P.; Barklage, A.; van Rooijen, B.; Seitz, A.; Horn, M.; Badrya, C.; Radespiel, R. Large-Scale Wind Tunnel Testing of an Advanced Hybrid Laminar Flow Control System. **2022**, 11 pages. https://doi.org/10.25967/570065.
18. Lecordix, J.L.; Mullender, A.; Lecossais, E.; Godard, J.L.; Hepperle, M. *Hybrid Laminar Flow Nacelle Design. ICAS* **1996** https://www.icas.org/icas_archive/ICAS1996/ICAS-96-2.3.4.pdf (accessed 2026-01-19).
19. Riedel, H.; Horstmann, K.-H.; Ronzheimer, A.; Sitzmann, M. Aerodynamic Design of a Natural Laminar Flow Nacelle and the Design Validation by Flight Testing. *Aerosp. Sci. Technol.* **1998**, *2* (1), 1–12. https://doi.org/10.1016/S0034-1223(98)80001-8.
20. APRE; CDTI Trl-Assessment-Tool-Guide-Final.Pdf. **2022** https://horizoneuropencpportal.eu/sites/default/files/2022-12/trl-assessment-tool-guide-final.pdf (accessed 2026-01-20).
21. Boeing Commercial Airplanes. 737 MAX by Design. **2017** https://www.boeing.com/commercial/ 737max/by-design (accessed 2026-05-13)
22. Wahlich, M.; Bismark, A.; Radestock, M.; Menchel, K.; Milot, M.; Cruz, D. Hybrid Laminar Flow Control Activities within the Frame of Clean Sky 2. In *AIAA SCITECH 2024 Forum*; American Institute of Aeronautics and Astronautics: Orlando, FL, **2024**. https://doi.org/10.2514/6.2024-1303.
23. Krishnan, K. S. G.; Bertram, O.; Seibel, O. Review of Hybrid Laminar Flow Control Systems. *Progress in Aerospace Sciences* **2017**, *93*, 24–52. https://doi.org/10.1016/j.paerosci.2017.05.005.
24. Young, T. M.; Fielding, J. P. *Flight Operational Assessment of Hybrid Laminar Flow Control (HLFC) Aircraft | Springer Nature Link*. **2001** https://doi.org/10.1007/978-3-540-45359-8_12.
25. Cholula, N. *NASA Tests Technology Offering Potential Fuel Savings for Commercial Aviation - NASA*. **2026** https://www.nasa.gov/centers-and-facilities/armstrong/nasa-tests-technology-offering-potential-fuel-savings-for-commercial-aviation/ (accessed 2026-05-10).
26. *Waypoint 2050 : Aviation: Benefits Beyond Borders. Air Transport Action Group* (ATAG) **2026** https://aviationbenefits.org/other-environmental-challenges/climate-action/waypoint-2050/ (accessed 2026-05-17).

Authors

Thomas is a high school student at St. Louis School, Milan. He is passionate about aviation and physics and hopes to combine these interests to pursue a degree in aeronautical engineering and a career in a related field.

DOI: 10.36838/IJHSR819.22

Appendix

Equation A1:

Using the formula for the laminar skin friction coefficient and using a value of 2 × 10^7:

$$Cf_{lam.} = \frac{1.328}{\sqrt{Re}}$$

$$Cf_{lam.} \approx 3 \times 10^{-4}$$

Equation A2:

Similarly, using the turbulent skin friction coefficient formula and Re value of 2 x 10^7 gives:

$$Cf_{turb.} = \frac{0.074}{Re^{1/5}}$$

$$Cf_{turb.} \approx 2.9 \times 10^{-3}$$

Equation A3:

Substituting previous values into dρ/dx gives:

$$\frac{\mathrm{d}\rho}{\mathrm{d}x} = \frac{1}{2}\,(0.38)\,(230)^2\,(-0.5) \approx -5000\,Pa/m$$



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

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