

AJUR

American Journal of
Undergraduate Research

Volume 20 | Issue 3 | December 2023

www.ajuronline.org

Print Edition ISSN 1536-4585
Online Edition ISSN 2375-8732

AJUR

American Journal of
Undergraduate Research

Volume 20 | Issue 3 | December 2023 | <https://doi.org/10.33697/ajur.2023.091>

- 2 **AJUR History and Editorial Board**
- 3 **Walking After Dark: A Sidewalk Illumination Case Study in Cedar City, UT**
Sabrina Waite, Ian Nelson, & Jamie Spinney
- 13 **Effect of Retinoic Acid on HaCaT and NIH-3T3 cells in an in vitro 3D Collagen Cell Culture Skin Model**
Olivia Atkins, Samantha Lopez, Yizhe Ma, & Louis Roberts
- 25 **Increasing Access to a Four-Year College: Impacts of a California State University Guaranteed Admission Program on College Enrollment Rates**
Joleen Chiu
- 51 **Conflicting Socio-Cultural Attitudes and Community Factors Resulting in Backstreet Abortion in Cato Manor, KwaZulu-Natal**
Chloe M. Sachs
- 69 **Dynamic Structural Equation Models: Promising Yet Concerning**
Suryadyuti Baral & Patrick J. Curran
- 81 **Finding the Fixing Number of Johnson Graphs $J(n, k)$ for $k \in \{2, 3\}$**
James Della-Giustina
- 91 **Anxiety as a Predictor of Movement During a Math Task in Elementary School Children**
Gabriella Snetkov, Anna E. Youngkin, Megan E. Lui, Analia Marzoratti, & Tanya M. Evans
- 99 **Chitosan-Silver Thin Film-Coated Titanium Coupons using Silane Linkers Inhibit Biofilm and Planktonic Growth**
Emily C. Montgomery, Nidhi Gupta, Matthew Atwill, Joel D. Bumgardner, & J. Amber Jennings
-

American Journal of Undergraduate Research (AJUR) is a national, independent, peer-reviewed, open-source, quarterly, multidisciplinary student research journal. Each manuscript of AJUR receives a DOI number. AJUR is archived by the US Library of Congress. AJUR was established in 2002, incorporated as a charitable not-for-profit organization in 2018. AJUR is indexed internationally by EBSCO and Crossref with ISSNs of 1536-4585 (print) and 2375-8732 (web).

EDITORIAL TEAM

Dr. Peter Newell, Editor-in-Chief
 Dr. Kestutis Bendinskas, Executive Editor
 Dr. Anthony Contento, Copy Editor

EDITORIAL BOARD *by subject area*

ACCOUNTING

Dr. Dean Crawford,
dean.crawford@oswego.edu

ARCHEOLOGY

Dr. Richard Redding,
rredding@umich.edu

ART HISTORY

Dr. Lisa Seppi,
lisa.seppi@oswego.edu

BEHAVIORAL NEUROSCIENCE

Dr. Aileen M. Bailey,
ambailey@smcm.edu

BIOCHEMISTRY

Dr. Kestutis Bendinskas,
kestutis.bendinskas@oswego.edu

Dr. Nin Dingra,
ndingra@alaska.edu

BIOENGINEERING

Dr. Jorge I. Rodriguez,
jorger@uga.edu

Dr. Jessica Amber Jennings,
jjennings@memphis.edu

BIOINFORMATICS

Dr. John R. Jungck,
jungck@udel.edu

Dr. Isabelle Bichindaritz,
ibichind@oswego.edu

BIOLOGY, PHYSIOLOGY

Dr. David Dunn,
david.dunn@oswego.edu

BIOLOGY, DEVELOPMENTAL

Dr. Poongodi Geetha-Loganathan,
p.geethaloganathan@oswego.edu

BIOLOGY, MICROBIOLOGY

Dr. Peter Newell,
peter.newell@oswego.edu

BOTANY

Dr. Julien Bachelier,
julien.bachelier@fu-berlin.de

CHEMISTRY

Dr. Alfredo Castro,
castroa@felician.edu

Dr. Charles Kriley,
ckriley@gcc.edu

Dr. Vadoud Niri,
vadoud.niri@oswego.edu

COMPUTER SCIENCES

Dr. Dele Oluwade,
deleoluwade@yahoo.com

Dr. Mais W Nijim,
Mais.Nijim@tamuk.edu

Dr. Bastian Tenbergen,
bastian.tenbergen@oswego.edu

COMPUTATIONAL CHEMISTRY

Dr. Alexander Soudackov,
alexander.soudackov@yale.edu

ECOLOGY

Dr. Chloe Lash,
CLash@stfrancis.edu

ECONOMICS

Dr. Elizabeth Schmitt,
elizabeth.schmitt@oswego.edu

EDUCATION

Dr. Charity Dacey,
cdacey@touro.edu

Dr. Marcia Burrell,
marcia.burrell@oswego.edu

EDUCATION, PHYSICS

Dr. Andrew D. Gavrin,
agavrin@iupui.edu

ENGINEERING, ELECTRICAL

Dr. Michael Omidiora,
momidior@bridgeport.edu

ENGINEERING, ENVIRONMENTAL

Dr. Félix L. Santiago-Collazo,
fsantiago@uga.edu

FILM AND MEDIA STUDIES

Dr. Lauren Steimer,
lsteimer@mailbox.sc.edu

GEOLOGY

Dr. Rachel Lee,
rachel.lee@oswego.edu

HISTORY

Dr. Richard Weyhing,
richard.weyhing@oswego.edu

Dr. Murat Yasar,
murat.yasar@oswego.edu

HONORARY EDITORIAL BOARD MEMBER

Dr. Lorrie Clemo,
lorrie.a.clemo@gmail.com

JURISPRUDENCE

Bill Wickard, Esq,
William.Wickard@KL.Gates.com

KINESIOLOGY

Dr. David Senchina,
david.senchina@drake.edu

LINGUISTICS

Dr. Taylor Miller,
taylor.miller@oswego.edu

LITERARY STUDIES

Dr. Melissa Ames,
mames@ein.edu

Dr. Douglas Guerra,
douglas.guerra@oswego.edu

MATHEMATICS

Dr. John Emert,
emert@bsu.edu

Dr. Jeffrey J. Boats,
boatsjj@udmercy.edu

Dr. Dele Oluwade,
deleoluwade@yahoo.com

Dr. Christopher Baltus,
christopher.baltus@oswego.edu

Dr. Mark Baker,
mark.baker@oswego.edu

MEDICAL SCIENCES

Dr. Thomas Mahl,
Thomas.Mahl@va.gov

Dr. Jessica Amber Jennings,
jjennings@memphis.edu

METEOROLOGY

Dr. Steven Skubis,
steven.skubis@oswego.edu

NANOSCIENCE AND CHEMISTRY

Dr. Gary Baker,
bakergar@missouri.edu

PHYSICS

Dr. Priyanka Rupasinghe,
priyanka.rupasinghe@oswego.edu

POLITICAL SCIENCE

Dr. Kaden Paulson-Smith,
Paulsonk@unwg.edu

Dr. Katia Levintova,
levintoe@unwg.edu

PSYCHOLOGY

Dr. Matthew Dykas,
mat.dykas@oswego.edu

Dr. Damian Kelty-Stephen,
keltystd@newpaltz.edu

Dr. Kenneth Barideaux Jr.,
kbarideaux@uscupstate.edu

SOCIAL SCIENCES

Dr. Rena Zito,
rzito@elon.edu

Dr. Dana Atwood,
atwoodd@unwg.edu

STATISTICS

Dr. Mark Baker,
mark.baker@oswego.edu

TECHNOLOGY, ENGINEERING

Dr. Reg Peccen,
regpeccen@sbsu.edu

ZOOLOGY

Dr. Chloe Lash,
CLash@stfrancis.edu

Walking After Dark: A Sidewalk Illumination Case Study in Cedar City, UT

Sabrina Waite ^{a*}, Ian Nelson ^b, & Jamie Spinney ^b

^a Department of Humanities and Social Sciences, Southern Utah University, Cedar City, UT

^b Department of Geosciences, Southern Utah University, Cedar City, UT

<https://doi.org/10.33697/ajur.2023.092>

Students: sabrinawaite123@gmail.com*, ian.andrew@comcast.net

Mentor: jamie.spinney@suu.edu

ABSTRACT

Sidewalks are an important part of public space, but they are becoming increasingly dangerous, especially at night. Therefore, it is important that sidewalks are adequately illuminated to meet the safety, comfort, and accessibility needs (*i.e.*, provide high levels of pedestrian reassurance) to sidewalk users. The objective of this case study was to quantitatively measure and subsequently explore the spatial and statistical distributions of light intensities along the sidewalks that surround Southern Utah University in Cedar City, Utah. A secondary objective was to develop a set of “adequate lighting” guidelines (*i.e.*, 5 to 10 lux) to help identify areas that could benefit from sidewalks lighting enhancements. A digital light meter was used to measure horizontal light intensity, in lux, at a systematic sample of 208 points along the sidewalks within the study area. Geographic coordinates were also collected for each sample observation to enable mapping and spatial analysis of the light intensity data. Results indicate 61% of the observations have lux values equal to zero (complete darkness), 13.4% met or exceeded the minimum guidelines, while 4.3% fell within the “adequate lighting” guidelines adopted by this case study. These results suggest that sidewalks are too dark in too many places to provide reassurance to sidewalk users in the study area. The mapping and spatial analysis results from this case study also provide information to support targeted visibility enhancements of sidewalks within the study area.

KEY WORDS

Lighting; Optimum Illuminance; Luminometer; Sidewalks; Pedestrians; Safety; Security; Reassurance

INTRODUCTION

Sidewalks are an important component of public space because they enable personal mobility and provide public access to goods and services that are required to meet one’s daily needs. Therefore, it is important for sidewalks to be safe, accessible, and well-maintained to enhance public health and maximize social capital.¹⁻³ This is especially true for people who are living with disabilities,⁴ but also for those people who do not or cannot drive a car, such as many older adults and all children.⁵ Sidewalk safety is a multidimensional concept that includes safety from potential trip hazards, pedestrian-vehicle crash hazards, and the hazards of personal crime. These, and other, considerations take on a special meaning at night, when visibility of sidewalk users is significantly reduced. Public space lighting (PSL), however, has been shown to enhance users’ perception of safety, comfort, and accessibility⁶⁻⁹ Adequate PSL levels are required at night for the safety, comfort, and accessibility of pedestrians, cyclists, and other sidewalk users by enabling people to detect potential hazards and then take appropriate action to avoid the hazard. Among these potential sidewalk hazards are “trip hazards”,¹⁰ which are defined by the Americans with Disabilities Act (ADA) as any vertical displacement of more than one-quarter inch¹¹ and may include ditches, excessive cross slopes,¹² or any abrupt vertical change in elevation caused by heaved slabs, tree roots, spalling, or cracks.¹³ At night, there is an increased risk of trips and falls when pedestrians are unable to detect unexpected obstacles or impediments along sidewalks.¹³ Therefore, adequate PSL is an important component of effective sidewalk design to ensure safe navigation at night for a wide range of users, their travel modes, and their physical abilities and preferences.

Safe navigation of the pedestrian environment at night must also consider the potential hazards associated with motor vehicle interactions. In the United States, more than 7,000 pedestrians were killed in traffic crashes with motor vehicles in 2020.¹⁴ Moreover, because poor lighting conditions reduce the visibility of pedestrians, almost three-quarters of these pedestrian deaths occurred at night.^{15, 16} This is especially compelling given that only about one-quarter of daily traffic volume occurs after dark.^{17, 18} However, the increasing popularity of automobiles, which started around the 1920s,¹⁹ further necessitates consideration of adequate levels of PSL for the pedestrian environment. Adequate levels of PSL are needed, because research indicates that road segments with a lower level of PSL tend to be associated with a higher number of pedestrian-vehicle crashes.²⁰ Conversely, research has shown that increased lighting levels have significantly reduced the number of pedestrian-vehicle crashes.²¹ While adequate PSL of sidewalks is important for the safety and security of people, property, and public spaces,^{7, 22} the design of our

pedestrian environment, and especially crosswalks, requires that pedestrians, cyclists, and other mobility device users (e.g., wheelchairs) are more visible to drivers,²³ regardless of the time of day.

In addition to trip hazards and traffic accidents, sidewalk users inherently feel less safe and less comfortable after dark. This lack of reassurance in one's safety after dark tends to be associated with decreased visibility of one's surroundings, increased opportunities for concealment, and typically fewer eyes on the street.^{6, 7, 24-26} There is a growing body of research that provides convincing evidence of increased PSL of sidewalks helps users feel safer (i.e., reassured) when walking at night.^{24, 27-29} The concept of "pedestrian reassurance"²⁷ has been described as the feelings of sidewalk users' safety, comfort, and confidence when walking alone after dark.^{30, 31} Designing sidewalks to promote pedestrian reassurance can collectively lead to a host of social, economic, environmental, plus personal, and public health benefits.⁷

In the context of pedestrian reassurance research, the definition of adequate lighting has been informed by the concept of an "optimum illuminance". Optimum illuminance is the threshold "beyond which further increase in illuminance has no significant effect on reassurance".²⁷ Previous studies into optimum illuminance found the spectrum of lighting intensity is strongly associated the spectrum of pedestrian reassurance.^{8, 27, 32-35} Moreover, the statistical distribution of the relationship between the two exhibits an area with rapid change (i.e., the escarpment) and another area beyond which there is negligible change (i.e., the plateau, or optimal level).^{8, 36} Despite a growing body of literature on the topic, research has not yet discovered a "precise estimate of optimal illuminance" (p. 1),³⁶ because the expectation of discovering a "one-size-fits-all" solution for nighttime illumination requirements is out of reach,⁸ considering the wide variety of contextual settings and different sidewalk users. Nevertheless, research by Svehkina et al.⁸ concluded that PSL of sidewalks in the range of 5 lx to 10 lx provided reasonably high levels of pedestrian reassurance, with only a negligible increase in reassurance with increasing illuminance. Similarly, this case study has adopted a set of adequate lighting guidelines to identify areas with sidewalks that may benefit from lighting enhancements. Following guidance from Svehkina et al.,⁸ other academic studies, and published guidelines^A, these guidelines are based on the breakpoints in a typical statistical relationship between illuminance and pedestrian reassurance with minimum and optimum thresholds of 5 and 10 horizontal lx, respectively. These guidelines were developed to provide guidance on the 'Goldilocks' locations and values that are considered adequately illuminated (i.e., just right), as well as the locations and values that are considered either too bright or too dark.

METHODS AND PROCEDURES

PSL consists of many different elements, such as intensity, glare, color temperature, and others. This case study examined light intensity^B (i.e., illuminance), which is defined as the density of incident luminous flux with respect to area at a point on a real or imaginary surface, and it can be measured on horizontal, vertical, or semi-cylindrical surfaces.³⁷ This case study employed a technology-based approach to measure light intensity on a horizontal plane using a digital light meter (i.e., Klein Tools, ET130), also known as a luminometer or luxmeter. While light intensity can be measured in foot candles (Imperial) or lux (metric), this case study used lux, where 1 lux equals the light output produced by a standard candle. For context, zero lux (lx) means the absence of detectable light (i.e., complete darkness), 0.1 to 0.2 lx is typical from a full moon on a clear night, 300-500 lx is typical for indoor light intensities, and about 1,500 to 100,000 lx is typical for outdoors on a cloudy versus a clear sunny day, respectively. The Klein Tools, ET130 can measure luminosities from 0 to 40,000 lux with a resolution of 0.1 lux and repeatability of +/- three percent. Despite the typically minimal impact of moonlight, this case study collected light intensity measurements within three days of the new moon phase in October 2022 (i.e., between the 22nd and 28th) to minimize the impact of moonlight.

With an interstate highway to the west, the study area for this project was a two-block neighborhood to the north, east, and south^C of the Southern Utah University (SUU) campus, which is a public university with a student population of 12,582.³⁹ SUU is located within Cedar City, which is a micropolitan area in the southwest part of Utah (see Figure 1 inset map) with a population of 35,235.⁴⁰ Similar to a similar recent study,⁸ this case study collected light intensity (lx) measurements at a systematic sample of points (i.e., about every 80 steps) along the sidewalks within the study area. The light intensity values (lx) from the luxmeter were

^A Similar research used segmented regression to identify discrete breakpoints in the relationship between illuminance and reassurance (i.e., feeling very unsafe to feeling very safe) and it found the plateau (i.e., optimum illuminance) ranged from 8.9 lx to 26 lx, depending on the location.³⁶ In line with the range proposed by Svehkina et al.,⁸ and specifically for sidewalks and walkways with high traffic (i.e., > 100 persons/ hour), the FHWA²⁰ recommends 10 semi-cylindrical lux for sidewalks as the optimum illuminance (i.e., additional light beyond this level does not increase visibility). At areas with higher likelihood of pedestrian-vehicle crashes, the FHWA recommends an average vertical illuminance of 20 vertical lux for midblock crosswalks and, conservatively, 30 vertical lux at intersections should provide adequate visibility for detection in most circumstances. Although not specific to sidewalks, but to provide context, subjective perceptions of safety, comfort, and visibility suggest an optimum illuminance of 10 horizontal lux in a parking garage and only 2 horizontal lux in parking lots.³⁸

^B Light intensity and illuminance are used interchangeably in this manuscript, but they are technically not equivalent. This case study used a digital light meter, or luxmeter, to take single point measurements of light intensity on a horizontal plane. The differences between illuminance and light intensity should not impact the conclusions of this study.

^C The study area did not include the area to the west of SUU campus, because the football field and facilities border Interstate 15 (I-15).

half (*i.e.*, 8) of them were associated with supplemental illumination sources, such as residential properties (n=2), businesses (n=2), commercial signs (n=1), university sports fields (n=2), or tennis courts (n=1).

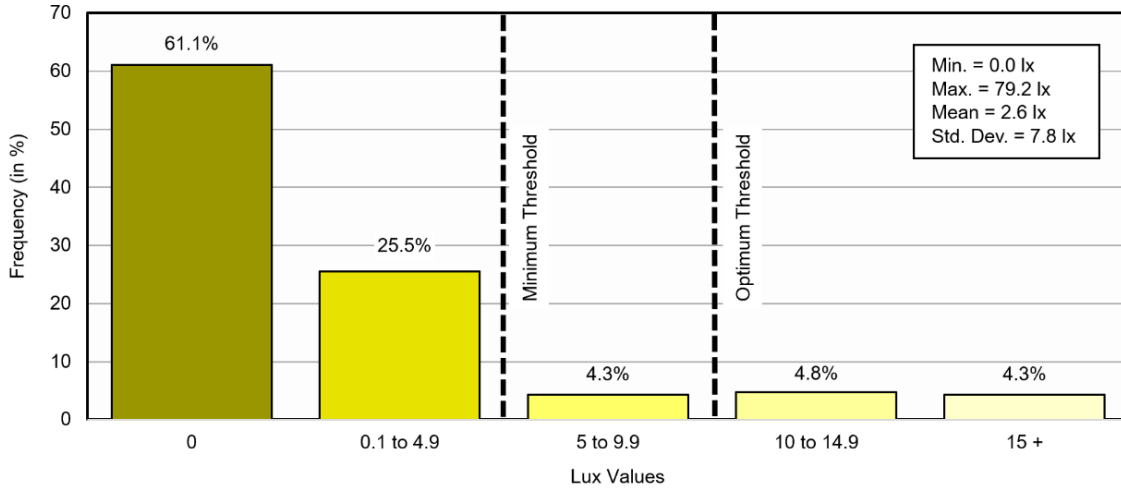


Figure 2. Frequency distribution of, and summary statistics for, sampled sidewalk light intensities (lx).

The map in Figure 3 uses a dark background of the street network with SUU campus located in the center to illustrate the spatial distribution of sampled sidewalk light intensities (lx) within the study area. The map uses graduated symbols with graduated color intensity to illustrate increasing categories of light intensity values, with the middle category illustrating the light intensities (5.0 – 9.9 lx) that met the adequate lighting guidelines adopted herein.

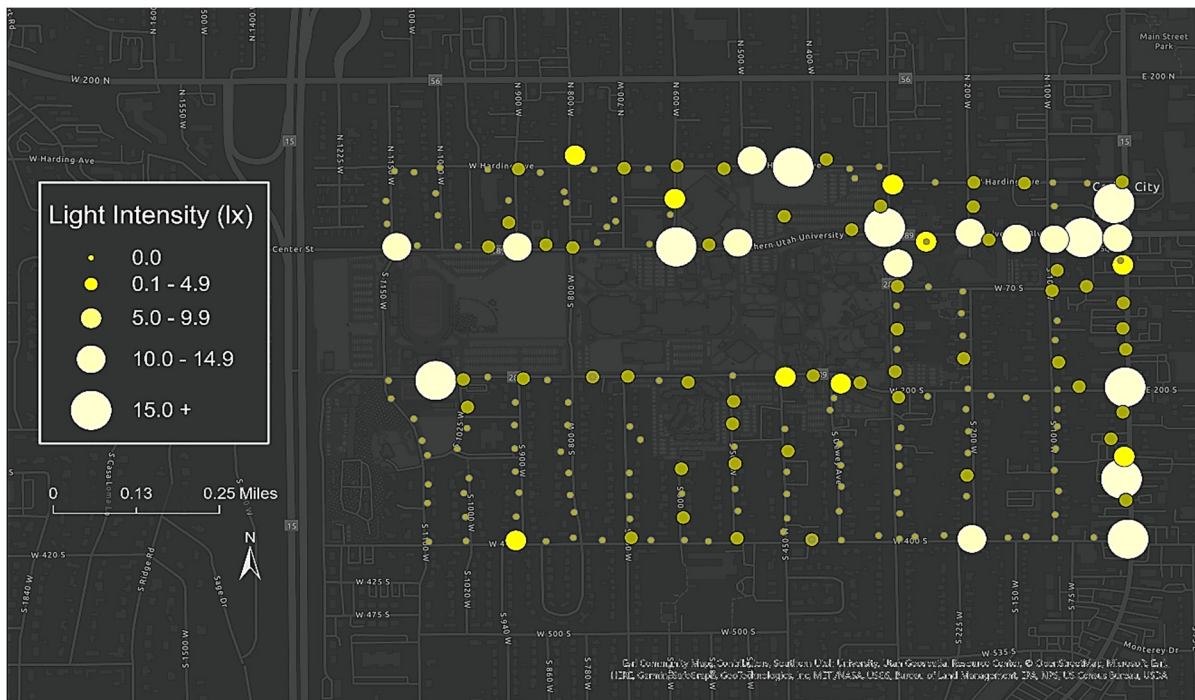


Figure 3. Spatial distribution of sampled and categorized light intensities (lx) within the study area.

Visual inspection of the light intensity (lx) values in Figure 3 suggests a clustered spatial pattern of high values on the northern side, and especially the northeast corner, of the study area with a smaller cluster of high light intensities in the southeast corner, which are both associated with commercial areas that border a state highway (a.k.a. Main Street). The study area also appears to have a significant cluster of low values, especially in the region south of SUU campus between W200 S and W 400 S. Otherwise, there appears to be random distribution of light intensity (lx) values across the study area. To quantify, and validate visual inspection, results from the Getis-Ord General G clustering statistic (z-score = 1.12, p = 0.262) suggest the light intensity (lx)

values are randomly distributed across the study area at a global-scale. However, results of the local-scale cluster and outlier analysis are illustrated in Figure 4, and they indicate several areas with significant local spatial clusters of apparent similarity (*i.e.*, high-high, or low-low values) as well as other clusters of apparent dissimilarity (*i.e.*, high value outliers surrounded by mostly low values, and low values surrounded by mostly high values).



Figure 4. Cluster and outlier analysis of sampled light intensities (lx) within the study area.

Figure 4 illustrates the results from cluster and outlier analysis, and the results confirm, from Figure 3, the presence of statistically significant clusters of high-high values in the northeast corner of the study area. The cluster and outlier analysis also identified the southwest corner as having the largest cluster of significant low-low values, meaning it is the least illuminated (*i.e.*, darkest) district within the study area. There is another, much smaller, cluster of low-low values in the northwest portion of the study area. The results in Figure 4 also illustrate several significant outliers of high values (*i.e.*, low-high outliers) amidst mostly low (*i.e.*, zero-lux) observations in the southeast corner of the study area and along S 200 W. Meanwhile, the outliers of low values (*i.e.*, low-high outliers) amidst mostly high values are in the east half of the study area and mainly along Highway 15 (aka. Main Street) in its northeast (*i.e.*, historic downtown) and southeast corners (*i.e.*, where Main St. intersects W 400 S).

DISCUSSION

Sidewalks provide public access to goods and services that people require to meet their daily needs. Unfortunately, the planning and design of sidewalks and the concern for pedestrian safety have been ignored too often and overlooked by too many.⁴⁴ This neglect is surprising given that community *walkability* (*i.e.*, the ability to safely walk to desired destinations within a reasonable distance) has several important benefits; it helps to reduce air pollution and greenhouse gas emissions, improve mental health, reduce chronic disease, foster social interaction, and enhance sense of place, among several other social, economic, and environment benefits.⁴³ Walkability is also an important measure of a community's livability and quality of life.⁴⁵⁻⁴⁷ Levels of sidewalk PSL directly impact walkability after dark by impacting actual and perceived levels of pedestrian safety, comfort, and security (*i.e.*, reassurance). Beyond perceptions of safety, statistics clearly indicate that pedestrians are the most vulnerable population of road users, especially at night. More specifically, pedestrians are almost seven times more vulnerable in the dark compared to daylight.⁴⁸ Given these statistics, the Federal Highway Administration argues that pedestrians have the potential to gain significant safety performance benefits from new or improved PSL.²⁰

In addition to safety from trip hazards and traffic accidents, PSL impacts perceived levels of safety and security (*i.e.*, pedestrian reassurance) in an area. Granted, PSL is only one of many factors affecting objective and perceived safety of urban public spaces after dark,^{7,49} but research suggests that approximately 40 percent of outdoor crimes happen in areas where illuminance values are 5 lx or darker compared to only 3 percent of outdoor crimes happen in areas where illuminance values are 20 lx or more.⁵⁰

Basically, when urban environments have adequate lighting, there tends to be an overall decrease in many types of deviant behavior.⁵¹ In fact, studies have examined the relationship between perceived safety and PSL, linking higher amounts of illuminance with a greater feelings of safety (FoS).^{8, 52} So, how much PSL is enough to provide adequate, or optimal, light intensity for effective planning and design of public spaces, including sidewalks, that would enable pedestrians, and drivers, to see potential hazards and meet the needs of most sidewalk users? Is more lighting automatically better, and for whom? Any discussion of lighting limits, especially if it is solely based on one dimension of lighting (*e.g.*, light intensity), must recognize the different lighting needs of many users, each with different preferences, abilities (*e.g.*, declining eyesight with age), and mobility devices (*e.g.*, cane, walker, scooter, power scooter, skateboard).

The definition of adequate lighting tends to be highly subjective, so any discussion of adequate lighting, or optimum illuminance, must recognize the importance of site- and situation-specific contextual factors.³⁶ Therefore, the practice of evaluating adequate lighting needs for pedestrian lighting varies widely among different regions, State Departments of Transportation (DOTs), and local agencies, and decisions are often made on a case-by-case basis.²⁰ Due to a lack of municipal- and state-wide standards, this case study adopted a set of objectively and statistically derived adequate lighting guidelines with lower and upper thresholds of 5 lx and 10 lx, respectively. While these guidelines are unlikely to meet the needs of all users, they were designed to assess whether the sidewalks in the study area are too dark, too bright, or adequate to promote pedestrian reassurance.

The results highlight especially low light intensities and poorly lit areas that are distributed across the study area. More specifically, only 4 percent of the sample observations fell within the adequate lighting guidelines and another 9 percent exceeded the guidelines. So, 13 percent of the sample observations either met or exceeded the minimum threshold of 5 lx that was adopted for this case study. That means 87 of the sample observations were too dark to provide pedestrians with feelings of safety, comfort, and security (*i.e.*, reassurance). It is important to emphasize, however, that light intensity is only one aspect of PSL that affects perceptions of safety, comfort, and security. Other dimensions include color temperature, glare, visual comfort, and a host of personal factors that impact individual FoS.²² For example, studies suggest that FoS and comfort increase in areas illuminated with white light compared to the same area illuminated with yellowish light.³⁰ While light intensity (*i.e.*, brightness) certainly has an objective dimension, it also has a subjective dimension that varies among individuals. Therefore, future research should follow the guidance of recent studies^{7, 8} that used smartphone apps to prompt participants to record their FoS in random locations under different lighting conditions to help determine subjective adequate lighting guidelines for the study area. Future research should also consider collecting pedestrian, or sidewalk user, counts to better identify (or prioritize) those areas that could most benefit from sidewalk lighting improvements.

This case study's limited sample size, straightforward research design, and emphasis on the objective dimension of light intensity (lx) means that this research is not without its limitations. For example, the time of year chosen for data collection impacted the results of this case study, because a few streetlights were burnt out, while other luminaires were malfunctioning and measured less than 1 lux. Consequently, the results of this case study only represent a limited temporal scale regarding sidewalk light intensities in the study area. For example, the data were collected during the autumn months, which resulted in a portion of the street trees retaining partial foliage that has the potential to reduce sidewalk illumination. On the other hand, leaf-off conditions (and the reflectivity of snow-covered surfaces) during the winter may have increased sidewalk illumination. The impact of street trees, and their foliage, was observed in at least two of the sample observations, where the trees blocked the light and created shadows that likely had a significant impact on measured sidewalk light intensity (lx). Similarly, another study noted that the type of tree and its maintenance along with luminaire design, type, height, and design may significantly affect the potential for street trees to impact sidewalk illumination.⁵³

The apparent local clustering of high sidewalk illumination values along the eastern edge of the study area (*i.e.*, Main Street) was similarly reported by a previous study⁵⁴ that found a strong association between nighttime illumination and land use. For example, commercial areas tend to have more illuminated signs, industrial areas tend to use artificial lighting for work activities, while parks and residential areas typically have lower levels of nighttime illumination. The relationship between nighttime illumination and land use can provide insight into the patterns of human activity and development in an area, which has important implications for planning the neighborhood lightscape. This land use-illumination relationship also raises some concerns associated with the proximity of different land uses, such as commercial facilities in residential areas, and the associated risk of increased exposure to light pollution.⁵⁴ Light pollution refers to the excessive or misdirected artificial light that interferes with the natural darkness of the night sky. This type of pollution can have numerous negative impacts on human health, wildlife, and the environment. For humans, exposure to artificial light at night can disrupt circadian rhythms, leading to sleep disorders, depression, and other health problems.⁵⁵ It can also interfere with stargazing and astronomy, thus hindering scientific research and cultural traditions.⁵⁶ Wildlife is also impacted by light pollution⁵⁷ particularly nocturnal animals such as birds, insects, and sea turtles. Furthermore, light pollution can also waste energy and contribute to carbon emissions, leading to climate change.⁵⁸ However, increased illuminance of the pedestrian environment does not need to contribute to light pollution. To mitigate these impacts, efforts are being made to

reduce the use of excessive lighting and ensure that outdoor light fixtures are fully shielded and directed downwards to illuminate only the pedestrian environment, to minimize light trespass and sky glow. These fixtures could also be installed on shorter light poles,⁵⁹ or install light fixtures in front of the crosswalk to increase the positive contrast of the pedestrian.⁶⁰

CONCLUSION

Safe, accessible, and well-maintained sidewalks are an important part of a community because they enable access to goods and services that people require to meet their daily needs. This is especially true for those people (including university students) who do not have access to a motor vehicle. It is also a timely discussion as urban planners aim to encourage more people to choose active modes of transportation (e.g., active communities, sustainable communities, healthy communities, 10-minute neighborhoods). However, inadequate sidewalk lighting poses a significant hazard to the safety, comfort, and security of pedestrians, which subsequently acts as a deterrent to choosing active modes of transportation after dark. The purpose of this case study was to quantitatively measure and analyze horizontal light intensity on sidewalks to determine where the light intensity was “adequate” and identify areas that are too dark (i.e., could benefit from lighting enhancements). Unfortunately, the results of this case study indicate that most sidewalks in the study area were inadequate; most areas are too dark and, consequently, failed to comply with this case study’s adequate lighting guidelines. The results from this case study strongly suggest that the sidewalks surrounding SUU’s campus are too dark to provide for pedestrians’, and other users’, safety, comfort, and security needs. Therefore, it is important to consider targeted visibility enhancements that increase illumination of sidewalks while ensuring appropriate design of these enhancements, so they do not contribute to light pollution.

REFERENCES

- Baobeid, A., Koç, M., and Al-Ghamdi, S. G. (2021) Walkability and its relationships with health, sustainability, and livability: Elements of physical environment and evaluation frameworks. *Front. Built Environ.* 7. <https://doi.org/10.3389/fbuil.2021.721218>
- National Association of City Transportation Officials (2013) Urban Street Design Guide. 3rd edition. Island Press: Washington, DC. <https://doi.org/10.5822/978-1-61091-534-2>
- Rajae, M., Echeverri, B., Zuchowicz, Z., Wiltfang, K., and Lucarelli, J. F. (2021) Socioeconomic and racial disparities of sidewalk quality in a traditional rust belt city. *SJM Popul Health* 9(16), 100975. <https://doi.org/10.1016/j.ssmph.2021.100975>
- Odame, P. K., and Amoako-Sakyi, R. O. (2020) Sidewalk accessibility and pedestrian safety among students with physical disability in the University of Cape Coast. *Curr. Res. J. Soc. Sci. and Humanities* 2(2), 109–122. <https://doi.org/10.3389/fbuil.2021.721218>
- Federal Highway Administration (2010) Safety Benefits of Walkways, Sidewalks, and Paved Shoulders. Washington, DC. http://safety.fhwa.dot.gov/ped_bike/tools_solve/walkways_brochure/walkways_brochure.pdf (accessed Aug 2023)
- Peña-García, A., Hurtado, A., and Aguilar-Luzon, M. C. (2015) Impact of public lighting on pedestrians’ perception of safety and well-being. *Saf. Sci.*, 78, 142–148. <https://doi.org/10.1016/j.ssci.2015.04.009>
- Portnov, B. A., Saad, R., Trop, T., Kliger, D., and Svechkina, A. (2020) Linking nighttime outdoor lighting attributes to pedestrians’ feeling of safety: An interactive survey approach. *PLoS ONE* 15(11). <https://doi.org/10.1371/journal.pone.0242172>
- Svechkina, A., Trop, T., and Portnov, B. A. (2020) How much lighting is required to feel safe when walking through the streets at night? *Sustainability*, 12(8), 3133. <https://doi.org/10.3390/su12083133>
- Trop, T., Shoshany Tavory, S., and Portnov, B. A. (2023) Factors affecting pedestrians’ perceptions of safety, comfort, and pleasantness induced by public space lighting: A systematic literature review. *Environ Behav*, 55(1–2), 3–46. <https://doi.org/10.1177/00139165231163550>
- Fotios, S., and Uttley, J. (2018) Illuminance required to detect a pavement obstacle of critical size. *Lighting Res Technol.* 50(3), 390–404. <https://doi.org/10.1177/1477153516659783>
- Americans with Disabilities Act of 1990, 42 U.S.C. § 12101 et seq. (1990). <https://www.ada.gov/pubs/adastatute08.htm> (accessed Apr 2023)
- Anderson, B., and Spinney, J. E. L. (2023) Evaluating driveway cross slopes and social equity in Cedar City, UT. *Am. J. Undergrad. Res.* 19(2), 3–9. <https://doi.org/10.33697/ajur.2022.059>
- Cohen, H. and Pauls, J. (2006) Warnings and Markings for Stairways and Pedestrian Terrain. In Wogalter M.S. (Ed.), *Handbook of Warnings*. CRC Press: Boca Raton, 711–722.
- Centers for Disease Control and Prevention. (2023) WISQARS (Web-based Injury Statistics Query and Reporting System) Atlanta, GA: US Department of Health and Human Services, CDC; 2020. <http://www.cdc.gov/injury/wisqars> (accessed Aug 2023)
- Hu, W., and Cicchino, J. B. (2018) An examination of the increases in pedestrian motor-vehicle crash fatalities during 2009–2016. *J Safety Res.* 67, 37–44. <https://doi.org/10.1016/j.jsr.2018.09.009>
- Ferenchak, N. N., and Abadi, M. G. (2021) Nighttime pedestrian fatalities: A comprehensive examination of infrastructure, user, vehicle, and situational factors. *J Safety Res* 79, 14–25. <https://doi.org/10.1016/j.jsr.2021.07.002>
- Griffith, M. S. (1994) Comparison of the safety of lighting options on urban freeways. *Public Roads*, 57(2). Federal Highway Administration, Washington, DC. <https://www.fhwa.dot.gov/publications/publicroads/94/fall/p94au8.cfm> (accessed Aug 2023)

18. CIE (2010) Lighting of Roads for Motor and Pedestrian Traffic [CIE 115:2010]. Commission Internationale de L'Eclairage, Vienna, Austria.
19. Boeschstein, W. (1985) Designing the nights of our cities. *J. Archit. Plan. Res.* 2(1), 55–64.
<https://www.jstor.org/stable/43028720>
20. Federal Highway Administration (2022) Pedestrian Lighting Primer. Virginia Tech Transportation Institute, and FHWA Office of Safety. Report No. FHWA-SA-21-087.
https://safety.fhwa.dot.gov/roadway_dept/night_visib/docs/Pedestrian_Lighting_Primer_Final.pdf (accessed Aug 2023)
21. Lutkevich, P., McLean, D., and Cheung, J. (2012) FHWA Lighting Handbook. Federal Highway Administration, Report No. FHWA-SA-11-22. https://highways.dot.gov/sites/fhwa.dot.gov/files/2022-09/fhwa_handbook2012.pdf (accessed Aug 2023)
22. Liu, M., Liu, Q., Li, Y., Liu, H., Jiao, W., Trop, T., Portnov, B. A., Liu, Y., Luo, T., and Zhang, B. (2022) Evaluating street lighting quality in residential areas by combining remote sensing tools and a survey on pedestrians' perceptions of safety and visual comfort. *Remote Sens.* 14(4), 826. <https://doi.org/10.3390/rs14040826>
23. Federal Highway Administration (2017) Crosswalk Visibility Enhancements, FHWA-SA-21-049.
https://highways.dot.gov/sites/fhwa.dot.gov/files/2022-08/PSC_New_Crosswalk%20Visibility%20Enhancements_508.pdf (accessed Aug 2023)
24. Boomsma, C., and Steg, L. (2014) Feeling safe in the dark: Examining the effect of entrapment, lighting levels, and gender on feelings of safety and lighting policy acceptability. *Environ Behav.* 46(2), 193–212. <https://doi.org/10.1177/0013916512453838>
25. Dravitzki, V. K., Cleland, B. S., Walton, D., and Laing, J. N. (2003) Measuring commuting pedestrians' concerns for personal safety and the influence of lighting on these concerns. In Proceedings of the 26th Australasian Transport Research Forum, Wellington, New Zealand, 1–3 October 2003.
26. Loewen, L. J., Steel, G. D., and Suedfeld, P. (1993) Perceived safety from crime in the urban environment. *J Environ Psychol.* 13, 323–331. [https://doi.org/10.1016/S0272-4944\(05\)80254-3](https://doi.org/10.1016/S0272-4944(05)80254-3)
27. Fotios, S., and Castleton, H. (2016) Specifying enough light to feel reassured on pedestrian footpaths, *LEUKOS*, 12(4), 235–243. <https://doi.org/10.1080/15502724.2016.1169931>
28. Vrij, A., and Winkel, F. W. (1991) Characteristics of the built environment and fear of crime: a research note on interventions in unsafe locations. *Deviant Behav.* 12, 203–215. <https://doi.org/10.1080/01639625.1991.9967873>
29. Blöbaum A., and Hunecke M. (2005) Perceived danger in urban public space: The impacts of physical features and personal factors. *Environ Behav.* 37, 465–486. <https://doi.org/10.1177/0013916504269643>
30. Knight, C. (2010) Field surveys investigating the effect of lamp spectrum on the perception of safety and comfort at night. *Lighting Res Technol.* 42(3), 313–330. <https://doi.org/10.1177/1477153510376794>
31. Fotios, S., Unwin, J., and Farrall, S. (2015) Road lighting and pedestrian reassurance after dark: A review. *Lighting Res Technol.* 47(4):449–469. <https://doi.org/10.1177/1477153514524587>
32. Wu, S., and Kim, M. (2006) The relationship between the pedestrians lighting environment and perceived safety. *J. Digit. Landsc. Archit.* 1, 57–66. https://gispoint.de/fileadmin/user_upload/paper_gis_open/DLA_2016/537612007.pdf (accessed Aug 2023)
33. Welsh, B. C., and Farrington, D. P. (2007) Improved Street Lighting and Crime Prevention: A Systematic Review. Stockholm, Sweden: Swedish National Council for Crime Prevention. <https://studylib.net/doc/13623708/improved-street-lighting-and-crime-prevention-a-systemati...> (accessed Aug 2023)
34. Boyce, P. R., Eklund, N. H., Hamilton, B. J., and Bruno, L. D. (2000) Perceptions of safety at night in different lighting conditions. *Lighting Res Technol.* 32(2), 79–91. <https://doi.org/10.1177/096032710003200205>
35. Fotios, S., Liachenko, M. A., and Uttley, J. (2019) Evaluation of pedestrian reassurance gained by higher illuminances in residential streets using the day-dark approach. *Lighting Res Technol.* 51(4), 557–575.
<https://doi.org/10.1177/1477153518775464>
36. Portnov, B., Fotios, S., Saad, R., and Kliger, D. (2022) Establishing optimal illuminance for pedestrian reassurance using segmented regression. *Lighting Res Technol.* 0(0). <https://doi.org/10.1177/14771535221080649>
37. CIE. (2020) CIE S 017/E:2020. ILV: International Lighting Vocabulary 2nd. Vienna (Austria): CIE. p. 248. <https://cie.co.at/e-ilv> (accessed Aug 2023).
38. Bhagavathula, R., and Gibbons, R. B. (2020) Light levels for parking facilities based on empirical evaluation of visual performance and user perceptions, *LEUKOS*, 16:2, 115–136. <https://doi.org/10.1080/15502724.2018.1551724>
39. National Center for Education Statistics. (2020) College Navigator – Southern Utah University.
<https://nces.ed.gov/collegenavigator/?q=southern+utah&s=all&id=230603#enrolmt> (accessed Aug 2023)
40. U.S. Census Bureau (2020) QuickFacts, Cedar City, Utah. <https://www.census.gov/quickfacts/cedarcitycityutah>. (accessed Aug 2023)
41. ESRI. (n.d.) An overview of the Spatial Statistics toolbox. Environmental Systems Research Institute; Redlands, CA, USA.
<https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/an-overview-of-the-spatial-statistics-toolbox.htm#:~:text=The%20tools%20in%20the%20Spatial,outliers%2C%20asses%20overall%20patterns%20of> (accessed Aug 2023)
42. ESRI. (n.d.) How High/Low Clustering (Getis-Ord General G) works. Environmental Systems Research Institute; Redlands, CA, USA. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/b-how-high-low-clustering-getis-ord-general-g-spat.htm> (accessed Aug 2023)

43. ESRI. (n.d.) Cluster and Outlier Analysis (Anselin Local Moran's I) (Spatial Statistics). Environmental Systems Research Institute; Redlands, CA, USA. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm> (accessed Aug 2023)
44. OECD. (1998) Safety of vulnerable road users. Scientific Expert Group on the Safety of Vulnerable Road Users (No. Report No. RS7). Paris, France: Organisation for Economic Co-operation and Development. https://safety.fhwa.dot.gov/ped_bike/docs/oecd_safety.pdf (accessed Aug 2023)
45. Jaśkiewicz M., and Besta, T. (2014) Is easy access related to better life? Walkability and overlapping of personal and communal identity as predictors of quality of life. *Appl Res Qual Life*. 9(3), 505–516. <https://doi.org/10.1007/s11482-013-9246-6>.
46. Mattson, J., Brooks, J., Godavarthy, R., Quadrifoglio, L., Jain, J., Simek, C., and Sener, I. (2021) Transportation, community quality of life, and life satisfaction in metro and non-metro areas of the United States. *Wellbeing, Space and Societ* 2, 100056. <https://doi.org/10.1016/j.wss.2021.100056>
47. Tobin, M., Hajna, S., Orychock, K., Ross, N., DeVries, M., Villeneuve, P. J., Frank, L. D., McCormack, G. R., Wasfi, R., Steinmetz-Wood, M., Gilliland, J., Booth, G. L., Winters, M., Kestens, Y., Manaugh, K., Rainham, D., Gauvin, L., Widener, M. J., Muhajarine, N., Luan, H., and Fuller, D. (2022) Rethinking walkability and developing a conceptual definition of active living environments to guide research and practice. *BMC Public Health* 22, 450. <https://doi.org/10.1186/s12889-022-12747-3>
48. Sullivan, J. M., and Flannigan, M. J. (1999) Assessing the Potential Benefit of Adaptive Headlighting Using Crash Databases [Report No. UMTRI-99-21]. University of Michigan Transportation Research Institute. <https://deepblue.lib.umich.edu/handle/2027.42/49401> (accessed Aug 2023)
49. Fisher, B. S., and Nasar, J. L. (1992) Fear of crime in relation to three exterior site features: Prospect, refuge, and escape. *Environ Behav*, 24(1), 35–65. <https://doi.org/10.1177/0013916592241002>
50. Stollard, P. (1991) *Crime Prevention Through Housing Design*. Taylor & Francis: London. <https://doi.org/10.4324/9780203223079>
51. Samuels, R. (2005) After-dark design, night animation, and interpersonal interaction: Toward a community-security paradigm. *J. Archit. Plan. Res.* 22(4), 305–318. <http://www.jstor.org/stable/43030749>
52. Zhang, J., and Dai, W. (2021) Research on night light comfort of pedestrian space in urban park. *Comput Math Methods Med.* <https://doi.org/10.1155/2021/3130747>
53. Hasson, P., Bradley, S., Walvatne, P., Lutkevich, P., and Leone, C. (2009) Trees, lighting, and safety in context-sensitive solutions. *Transp. Res. Rec.* 2120(1), 101–111. <https://doi.org/10.3141/2120-11>
54. Cheon, S., and Kim, J-A. (2020) Quantifying the influence of urban sources on night light emissions, *Landsc Urban Plan* 204. <https://doi.org/10.1016/j.landurbplan.2020.103936>.
55. Rodrigo-Comino, J., Seeling, S., Seeger, M. K., and Ries, J. B. (2021) Light pollution: A review of the scientific literature. *Anthr. Rev.* 0(0). <https://doi.org/10.1177/20530196211051209>
56. Kyba, C. C. M., Altintas, Y. O., Walker, C. E., and Newhouse, M. (2023) Citizen scientists report global rapid reductions in the visibility of stars from 2011 to 2022. *Science* 379(6629), 265–268. <https://doi.org/10.1126/science.abq778>
57. Raap, T., Pinxten, R., and Eens, M. (2015) Light pollution disrupts sleep in free-living animals. *Sci. Rep.* 5, 13557. <https://doi.org/10.1038/srep13557>
58. Tatro, K. (2020) Light energy: Our wasted resource. *Consilience*, (22), 65–72. <https://doi.org/10.7916/consilience.vi22.6731>
59. Cevallos, F. (2016) Safe and Accessible Pedestrian Facility Inventory Model (SAPFIM), National Center for Transit Research (NCTR) Report No. CUTR-NCTR-RR-2016-15, Center for Urban Transportation Research, University of South Florida. https://scholarcommons.usf.edu/cutr_nctr/21 (accessed Aug 2023)
60. Bhagavathula, R., and Gibbons, R. B. (2023) Lighting strategies to increase nighttime pedestrian visibility at midblock crosswalks. *Sustain.*, 15(2), 1455. <https://doi.org/10.3390/su15021455>

ABOUT STUDENT AUTHOR(S)

Sabrina Waite is a senior majoring in both History and French, and she plans to graduate in the spring of 2024. Ian Nelson is a sophomore majoring in Geospatial Science, and he plans to graduate in the spring of 2026.

PRESS SUMMARY

This case study measured light intensity (lx) on sidewalks using a digital light meter. The results showed that most of the study area was in complete darkness (*i.e.*, zero lx) and 87 percent of the sample observations were too dark to meet the safety, comfort, and accessibility needs of sidewalk users. The results also highlight areas that require enhancements to visibility along the sidewalks within the study area.

Effect of Retinoic Acid on HaCaT and NIH-3T3 cells in an *in vitro* 3D Collagen Cell Culture Skin Model

Olivia Atkins^{a1}, Samantha Lopez^{b1}, Yizhe Ma^a, & Louis Roberts^{*a}

^aDepartment of Biology & Biotechnology, Worcester Polytechnic Institute, Worcester, MA

^bDepartment of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA

¹Authors contributed equal intellectual effort to this manuscript; order determined alphabetically.

<https://doi.org/10.33697/ajur.2023.093>

Students: splopez@wpi.edu, omatkins@wpi.edu

Mentor: laroberts@wpi.edu*

ABSTRACT

Human skin aging is characterized by epidermal and dermal thinning, loss of elasticity, and wrinkles. Keratinocytes, the most common type of skin cell and fibroblasts, present in the cellular stroma beneath the skin's surface, each play a role in aging. Using these cell types in *in vitro* research can reveal a deeper understanding of the dermatological function and cellular changes in aged skin. 3D cell culture techniques provide an opportunity to use these cell types in a model that can more accurately mimic human skin. Treatment of aging skin is of interest to both medical and consumer communities. Retinoic acid (RA) is a metabolite of vitamin A and retinol that assists in cell proliferation, differentiation, and immune functions. Over the counter (OTC) and prescription retinoids are common topical products used for anti-aging and acne treatments. This study seeks to determine the impact of topical retinoid creams on keratinocyte proliferation and morphology in 3D cell culture models of aged and unaged human skin. NIH-3T3 fibroblasts were embedded in a 3D collagen matrix of varying thickness, and HaCaT keratinocytes were seeded on top of the matrix at varying seeding densities to mimic aging and youthful skin. 0.1 μ M 0.025% tretinoin and 0.1 μ M 0.1% adapalene topical creams were prepared in culture medium and used to treat cells daily, on alternate days, or just once during a week-long period. Alamar Blue assays and microscopy showed that tretinoin treatment was cytotoxic at this concentration, with a single treatment reducing cell viability by ~43% compared to the untreated control. Adapalene treatment, while showing significantly greater cell proliferation than tretinoin, did not exceed the proliferation of the untreated control. It is understood that retinol increases cell turnover by killing cells rapidly, so it is proposed that in our model, the rate of proliferation does not overcome the rate of cell death. Cell viability trends remain similar between young and old skin model treatments. Future studies should focus on creating a 3D model more accurately representing the aging *in vivo* skin environment where keratinocytes can be readily differentiated from epidermal stem cells.

KEYWORDS

HaCaTs; NIH-3T3; Retinoids; Retinoic Acid; Tretinoin; Adapalene; 3D Collagen Gel; Proliferation

INTRODUCTION

Epithelial cells are polygonal-shaped cells with regular, consistent dimensions. In two-dimensional culture, they grow in patches, form a monolayer, and tend not to migrate. Keratinocytes are the most common type of skin cell and function as the structural and barrier components of the epidermis, the outermost layer of the skin. Keratinocytes differentiate from epidermal stem cells and originate from the stratum basale, the deepest layer of the epidermis. As they age, grow, and differentiate into keratinocytes, they make their way to the epidermis, where they stay until they reach the final layer of the stratum corneum as corneocytes, where they die and are eventually shed by the body. Through differentiation, they produce keratin, the protein that strengthens the skin, hair, and nails.¹ These cells can be used to study the release of inflammatory and repair mediators. These cells can be used as a method to study the molecular mechanisms of epidermal regulation, homeostasis, pathophysiology, wound healing, and dermatologic diseases in *in vitro* models.³ HaCaT cells are a nontumorigenic monoclonal cell line of spontaneously immortalized keratinocytes derived from human skin. HaCaTs are epithelial in nature and fully differentiate under culture conditions.³ After stimulation from different calcium concentrations in the medium, HaCaT cells can form stratified layers, differentiate, and express differentiation markers. In addition, they can revert back and forth between differentiated and basal states.²

NIH-3T3 cells are fibroblasts derived from mouse embryos.⁴ Fibroblasts are a key component of the stroma and, more specifically, the extracellular matrix (ECM). In connective tissue stroma, fibroblasts make several important products, including collagen, glycans, and prostaglandins. Commonly characterized by their plasticity, fibroblasts can differentiate to have many distinct functions. In the skin, for example, superficial fibroblasts hold a role in follicle formation and wound healing, while deeper cells are involved in ECM reorganization.⁵ These cells are in constant communication with their surroundings and respond to both autocrine and paracrine signaling. Fibroblasts can interpret these signals and synthesize or remodel the ECM with regulation from various growth factors and pathways. Fibroblast structure is characterized by a star-like shape with several cytoplasmic projections.⁵ In culture, the live cells are adherent and migratory. This specific 3T3 cell line was established in the 1960s, and since then, it has advanced the scientific understanding of fibroblast biology and lineage.⁶ Fibroblasts are a valuable component of many cell culture models. In 3D models specifically, fibroblast-collagen matrices provide cell cultures that more closely represent the *in vivo* environment than traditional 2D cultures. Aspects such as cell-to-cell signaling and cell migration in a 3D fibroblast-collagen matrix more closely mimic those physiological conditions in the extracellular matrix.⁷

Collagen has a highly complex and hierarchical conformation with a recognizable triple helix secondary structure of amino acids.⁸ Collagen is the most abundant protein in animals and is the primary component of the extracellular matrix, with 90% of the 29 identified types of collagens in humans being fibrillar.⁹ Collagen I is the most common fibrillar collagen for cell culture, makes up the major structural component of many tissues, and can be found in skin, bone, tendons, and other connective tissues.⁹ Collagen is the most used naturally derived hydrogel and coating for tissue engineering.¹⁰ Collagen I is used in 2D cell culture as a coating or base and in 3D cell culture as a neutralized hydrogel to replicate the *in-vivo* environment of the body. 3D collagen hydrogels are used widely in cell culture, biotechnology, and medicine due to their polymer network and high-water content. As defined by Andersen et al., a three-dimensional cell culture involves cells embedded in a 3D matrix where cell and ECM signaling can be sent and received in all directions.¹¹ Cells can be embedded in the 3D collagen hydrogels or seeded on top of the hydrogel, serving as a 2D surface.⁶ For example, a commonly used concentration of collagen is 2 mg/mL, a low concentration, specifically used for embedding cells as the hydrogel is less stiff and gives cells higher motility and higher cell migration velocity.¹² Collagen is advantageous for cell culture because of its biometric cytocompatibility, ability to assist cell adhesion, promote cell growth and differentiation, and can provide a similar viscoelastic environment.⁷

Retinoic acid (RA) is a metabolite of vitamin A that assists in the vitamin's related biological activities. RA is also a natural metabolite of retinol that assists in cell proliferation, differentiation, and immune functions.¹³ Researchers have studied the growth and differentiation of human keratinocytes (HaCaTs) in response to RA *in vivo* and *in vitro* by evaluating differentiation through keratin synthesis and production; they found that RA was correlated to cell proliferation *in vitro*.¹⁴ In addition, they discovered that HaCaT cells expressed specific keratins dependent on RA and calcium levels in the media.¹⁴ When compared to untreated keratinocytes, HaCaTs treated with retinoic acid showed a mild increase in cell proliferation at certain doses.¹⁴

When skin ages, it undergoes structural changes such as thinning of its epidermal and dermal layers as well as weakening of the junction between these layers.¹⁵ In a study with women aged 18-69, it was found that there is progressive and significant thinning, 6.4% per decade, of the epidermis beginning at age 30.¹² Mimicking these changes *in vitro* can be achieved by obtaining skin explants from elderly donors or treating cells with age-inducing chemicals. Researchers have developed a 3D model of aging skin to replicate the phenotypic changes observed in aging human skin using a collagen-based matrix seeded with normal human fibroblasts and then seeded healthy human skin cells from young subjects on top.¹⁶ They found that in these prolonged cultures, epidermal thickness decreased over time.¹⁶ Further analysis also indicated similarities to naturally aging skin, such as the thickening of the lamina densa, a part of the skin between the epidermis and dermis.¹⁶ Another study used human skin models, to represent aging by manipulating the seeding concentration of senescent and normal fibroblasts in the matrix while monitoring keratinocyte proliferation and differentiation.¹⁷ In their aged models they used specific keratinocyte seeding densities and fibroblast densities to more accurately represent *in vivo* skin.

Human skin needs to be examined experimentally to understand how different types of retinoic acid help its rejuvenation. Investigating this topic using a 3D cell culture more accurately models the *in vivo* skin environment. Topical Vitamin A treatments, such as retinol, retinoid, tretinoin, and adapalene, are used to treat acne by unclogging pores and providing anti-aging effects by increasing collagen production and cell proliferation while exfoliating and removing dead skin cells.¹⁸ Tretinoin is a prescription retinoid whereas adapalene is sold over the counter. Since tretinoin is more concentrated it is used for different purposes like treatment of cystic acne and to assist in the healing of acne scars. Adapalene is used for skin rejuvenation, prevention and removal of wrinkles and fine lines.¹⁹ The concentration of each drug, adapalene and tretinoin, is determined by several factors such as potency, effectiveness, and safety profile. Tretinoin and adapalene's concentration are also related to their mechanisms of action. Tretinoin is a highly potent retinoid that binds to nuclear retinoic acid receptors and regulates gene expression and leads to changes in cellular differentiation, proliferation, and apoptosis.²⁰ The effectiveness of tretinoin is dose-dependent and higher concentrations of the drug can lead to greater therapeutic effects. However, tretinoin can also be highly irritating and cause side

effects such as redness, peeling, and dryness of the skin.²⁰ On the other hand, adapalene is a less potent retinoid that acts by binding to specific retinoic acid receptors in the skin and regulating sebocyte differentiation and proliferation.¹⁹ Adapalene is generally less irritating and handled better by users than tretinoin.¹⁹ Therefore, lower concentrations of adapalene can still achieve similar therapeutic effects while minimizing the harmful side effects.

Previous studies have investigated the molecular basis of retinoids *in vivo* for cosmetic purposes.²¹ Tretinoin and other FDA-approved retinoids are more thoroughly studied, with the mechanisms of action, genomic and nongenomic, well defined.²² They also have clinical evidence to support their efficacy.²² More must be understood about the comparative efficacy of over-the-counter (OTC) retinoids, like adapalene, which are readily available and popular for cosmetic use. As previously mentioned, retinoids show anti-aging effects, however, there need to be more studies surrounding these effects *in vitro*, specifically in 3D skin models. Our study begins to address this gap by creating a 3D model of human aging skin to investigate the proliferation rate of human epidermal tissue cells after treatment with prescription tretinoin and OTC adapalene creams. In this study, we show how HaCaTs (human epithelial cells), proliferate and behave in the presence of topical vitamin A treatments in *in vitro* 3D cell culture.

METHODS AND PROCEDURES

Cells and Culture Conditions

NIH-3T3 (ATCC, Catalog No. CRL-1658) and HaCaT (AddexBio, Catalog No. T0020001) cells were obtained. These cell lines were cultured at 37 °C and 5% CO₂ in complete medium that contains Dulbecco's Modified Eagle Medium (DMEM) (Corning, Catalog No. 10-013-CV) with 1x penicillin-streptomycin and 10% fetal bovine serum (FBS). Dulbecco's Modified Eagle Medium (DMEM) has a calcium concentration of 1.8 mM.²³ Cells were routinely passaged with 5% trypsin every two days prior to seeding in the 3D model.

3D Gel Model

To embed NIH-3T3 fibroblasts in a 3D collagen gel with a layer of HaCaTs on top of the gel a multi-step procedure was performed; a schematic can be found in **Figure 1** below. To mimic aged skin, the layer of fibroblasts embedded in collagen was 1 mm thick, as this was the smallest thickness that would cover a 24-well plate and be correlated to skin age. Since there is a 6.4% decrease in epidermis thickness every 10 years, younger skin was chosen to be 2 mm thick.²⁴ To seed 2 mm thick collagen gel, 0.2 mL of the fibroblast embedded collagen gel solution was needed to create this thickness, and 1 mm thick collagen gel required 0.1 mL of the fibroblast embedded collagen gel solution. The fibroblast-embedded collagen gel solution was 20% cell suspension and 80% 3D collagen. The gels were seeded with a fibroblast density of 0.55 x 10⁶ cells/mL. So, a T75 flask of confluent NIH3T3 cells, at a concentration of approximately 0.7x10⁶ cells/mL, was trypsinized with 1 mL of 5% trypsin and centrifuged for 5 minutes in a clinical centrifuge at 3000 rpm. The cell pellet was resuspended into 1 mL into complete media, counted, and resuspended again in 0.9 mL of complete media to create the 20% of 4.5 mL needed for the cell suspension. To make the fibroblast-embedded collagen gel solution, 3.6 mL of 3 mg/mL 3D collagen gel (Sigma Aldrich Catalog No. C4243) was mixed with 0.9 mL of the cell suspension was added (see Appendix). To add the control gels to the plate, 0.2 mL, or 0.1 mL of 3D collagen gel without fibroblast cells was added to non-fibroblast control wells. To add the remaining control and treatment gels to the plate, 0.2 mL or 0.1 mL of the fibroblast-embedded collagen gel solution was added to the wells. The plate was incubated for 1 hour for the gel to harden. A schematic of the plate can be found in **Table 1** below.

To mimic skin, a layer of HaCaTs were seeded on top of the fibroblast-embedded collagen gel with a density of 3500 cells/cm² to mimic aged skin and 7000 cells/cm² to mimic younger skin.¹⁷ The surface area of a 24-well plate is 1.9 cm² and it has a volume capacity 0.75 mL of complete media. The cell density needed for the aged skin model was 6650 cells per well with 0.75 mL of media, and 10 mL of this cell suspension was needed to add to the top of 13 wells; the math can be seen in **Equation 1** below. Therefore, the cell density for the younger skin model was 13,300 cells per well in 0.75 mL media. To seed this layer of HaCaTs on top of the fibroblast collagen gel, a confluent T75 flask of HaCaT cells was trypsinized with 1 mL of trypsin, resuspended in 1 mL of complete media, counted, and resuspended in the correct amount of media to make the needed cell densities for each treatment group. After the fibroblast-embedded collagen gel solidified, 0.75 mL of the correct cell suspensions were added to the respective wells according to the schematic in **Table 1**. Finally, the plate was incubated at 37 °C with 5% CO₂ in a humid environment.

3500 cells/cm² with 0.75 mL media
24 well plate = 1.9 cm² surface area

$$\frac{3500 \text{ cells}}{\text{cm}^2} \times 1.92 \text{ cm}^2 = 6.650 \text{ cells per well with 0.75 mL media}$$

$$\frac{6.650 \text{ cells}}{0.75 \text{ mL media}} \times \frac{\text{cells per 1 mL}}{1 \text{ mL media}} = 8867 \text{ cells/mL}$$

Equation 1.

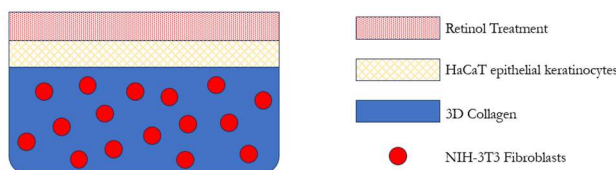


Figure 1. This Figure shows the collagen-cell matrix design. NIH-3T3 fibroblasts are embedded in a 3D collagen gel with a thickness of either 2 mm (0.2 mL) or 1 mm (0.1 mL) and HaCaT epithelial cells are seeded on top of the gel.

	1	2	3	4	5	6
A	2 mm CG + H + M	2 mm CG(F) + H + M	1 mm CG(F) + H + M	2 mm CG + M	1 mm CG + M	M only
B	2 mm CG(F) + H + T daily treatment	2 mm CG(F) + H + T alternating treatment	2 mm CG(F) + H + T single treatment	1 mm CG(F) + H + T daily treatment	1 mm CG(F) + H + T alternating treatment	1 mm CG(F) + H + T single treatment
C	2 mm CG(F) + H + A daily treatment			1 mm CG(F) + H + A daily treatment	1 mm CG(F) + H + A alternating treatment	1 mm CG(F) + H + A single treatment
D		2 mm CG(F) + H + A alternating treatment	2 mm CG(F) + H + A single treatment			

M = Complete DMEM
 CG = Collagen Gel
 CG(F) = Collagen Gel with NIH-3T3 fibroblast cells embedded
 H = HaCaT cells
 A = Adapalene treatment
 T = Tretinoin treatment

Table 1. Retinoid Treatment Plate Set up.

Retinoid-Media Preparation

Perrigo Tretinoin Cream, USP 0.025% and Differin Gel Adapalene 0.1% were used. Due to retinoid sensitivity to both light and heat, these sterilization methods were not possible without altering the potency of the creams. Tretinoin and adapalene creams were streaked on LB agar plates and incubated for two days at 37°C to assess sterility. No growth was visible on any of the plates. 1 μM retinoid-media solutions were prepared by first massing specified amounts of 0.025% tretinoin cream and 0.1% adapalene cream according to the calculations in **Equations 2, 3, and 4** below for a final media volume of 20 mL. **Equation 2** shows the calculation of the desired 1 μM concentration in moles. **Equation 3** calculates the mass required for 1 μM 0.025% Tretinoin in 20 mL media. **Equation 4** calculates the mass required for 1 μM 0.1% Adapalene in 20 mL media. After measuring the appropriate retinoid weight, the cream was resuspended in 20 mL complete DMEM, 10% FBS as used in culturing described above. The solutions were filter sterilized using a 50 mL Steriflip filter unit (Millipore Sigma, Catalog No. C3238). Once prepared, the retinoid-media solutions were stored at 4 °C and equilibrated to room temperature prior to cell treatment.

$$1 \mu\text{M} \times 20 \text{ mL} = 20 \times 10^{-9} \text{ mol} \tag{Equation 2.}$$

$$\frac{1 \mu\text{M} \text{ 0.025\% Tretinoin}}{(300.44 \text{ g/mol} / (20 \times 10^{-9} \text{ mol})) / 0.00025} = 2.4 \times 10^{-2} \text{ g in 20 mL media} \tag{Equation 3.}$$

$$\frac{1 \mu\text{M} \text{ 0.1\% Adapalene}}{(412.52 \text{ g/mol} / (20 \times 10^{-9} \text{ mol})) / 0.001} = 8.25 \times 10^{-3} \text{ g in 20 mL media} \tag{Equation 4.}$$

Retinoid Treatments

After the plate was prepared and incubated for the cells to reach the correct confluence, the wells were treated with the 1 μM adapalene and tretinoin media prepared above. DMEM was removed from the wells and replaced with the respective treatment media. In the original treatment plate, control models were treated with complete DMEM every other day. For testing the impact

of tretinoin on cells, young and old skin models were treated with either consecutive 5 days (days 1, 2, 3, 4, and 5), 3 alternating days (days 1, 3, and 5), or a single first day (day 1) treatment of 0.1 μ M 0.025% tretinoin. Young and old skin models were treated with 0.1 μ M 0.1% adapalene in the same manner. Each well was observed daily, with a final imaging and proliferation assay readout on the 9th day after treatment.

Microscopy

During treatment, each well was imaged every day the media was changed on days 1, 2, 3, 4, 7, and 9 with the Olympus IMT-2 inverted microscope at 100X magnification, under 10X objective with a 10X eyepiece. After treatment concluded, HaCaT cells were imaged using a Cytation Gen 5 microscope under brightfield at 10X objective magnification.

Alamar Blue Resazurin Proliferation Assay

To assess cellular activity dependent on retinoid treatment conditions, a cell proliferation assay was completed with Alamar Blue Cell Viability Reagent (Invitrogen, Catalog No. DAL1025). For T=0 data, wells with 2 mm collagen-cell model, 1 mm collagen-cell model, and media only were treated 72 hours after seeding with Alamar Blue reagent at 10% of the gel and media volume in the well. For example, the wells with 0.2 mL collagen gel and 0.75 mL of media had 95 μ L of Alamar Blue added. The plate was incubated at 37 $^{\circ}$ C and 5% CO₂ for 40 hours. The fluorescence of each well was measured using an excitation wavelength of 540 nm and an emission wavelength of 590 nm. This fluorescence intensity was measured in relative fluorescence units (RFU). Each measurement had an associated blank measurement without cells which was used to determine the baseline fluorescence of the media and collagen gel. The blank measurement value was subtracted from the cell dilution measurements to obtain fluorescence values from the cellular activity. This procedure was repeated 72 hours after completion of the treatment plate.

RESULTS

Proliferation Imaging Results

The wells were imaged every time the media were changed to observe HaCaT confluency and cell morphology. Models without cells were not imaged. Through visual observations, the HaCaTs proliferated normally in control models, treated with complete media. Controls included HaCaTs seeded on top of 2 mm thick 3D collagen gel with no fibroblasts embedded inside the gel and HaCaTs seeded on top of 2 mm and 1 mm thick 3D collagen gel with fibroblasts embedded inside the gel. Throughout the experiment, these models maintained a healthy cell morphology, with the cells being very confluent by the end of the experiment on day 9, as shown in **Figure 2**. As the experiment goes on the cells become more confluent as expected with very few floaters every day. The few floaters were removed with media changes every other day. The first day after seeding the HaCaTs seeded on top of the gel with fibroblasts embedded were more confluent than the HaCaTs seeded without fibroblasts embedded.

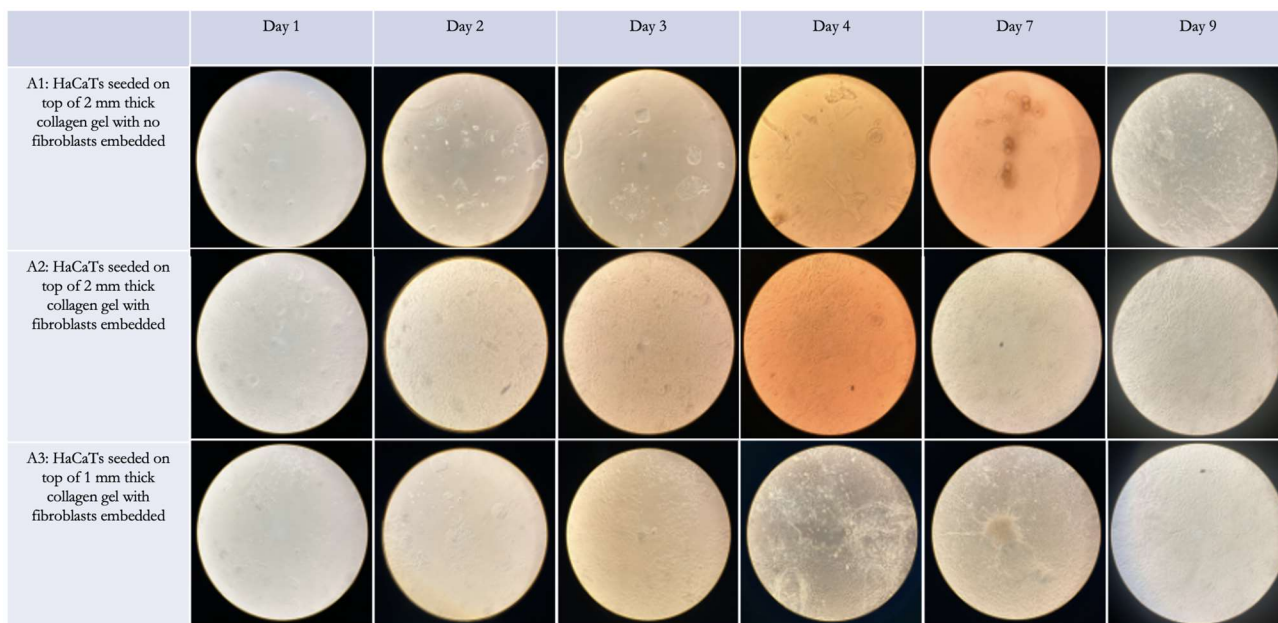


Figure 2. Daily imaging of HaCaT control wells A1 (2 mm collagen gel, no fibroblasts) and A2/A3 (2 mm and 1 mm collagen gel with fibroblasts) shows increasing confluency, reaching full confluence by day 9 with healthy cells.

The tretinoin treatments produced cells that balled up, detached from the gels, and floated. Images correlating to tretinoin treatment wells can be found in **Figures 3** and **Figure 4** for the younger and aged treatments respectively. Throughout the course

of the tretinoin treatment, the wells with less treatment frequency had healthier-looking and more confluent cells. The wells treated with tretinoin every day were losing cells when the media was replaced; the remaining cells were circular and not attached to the gel and by day 9 the well had lost most of its cells and was not very confluent; this may be due to tretinoin’s mechanisms of action.

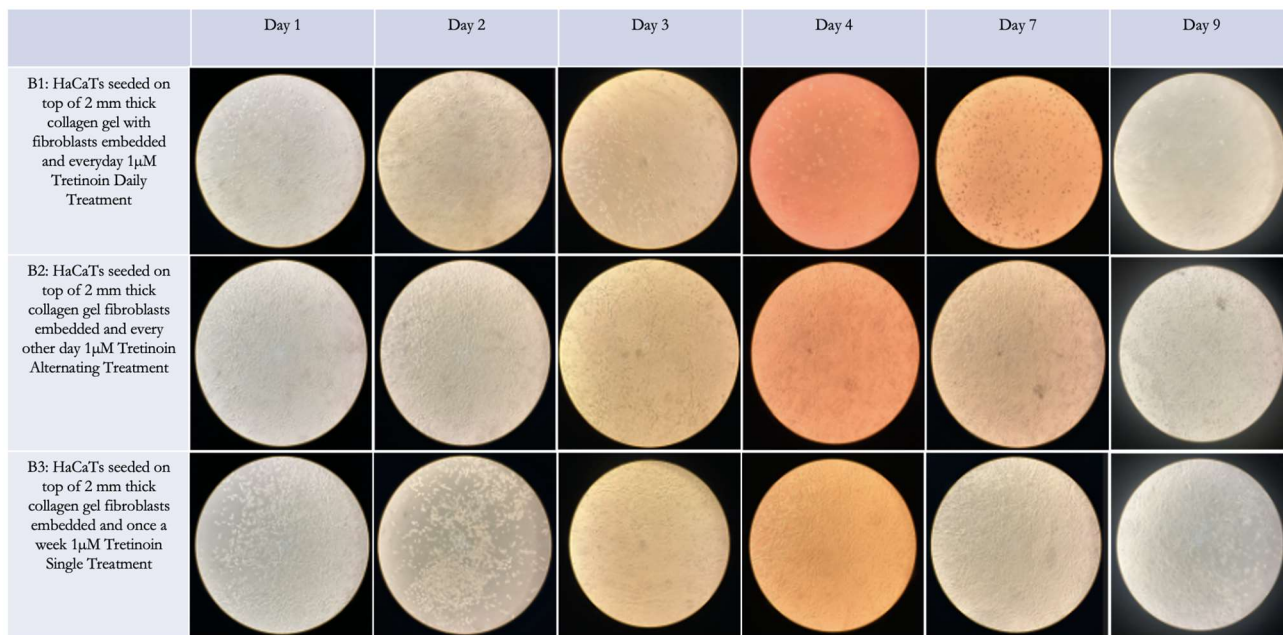


Figure 3. Daily imaging of HaCaT cells on 2 mm collagen gel with embedded fibroblasts. B1 received daily 1 µM tretinoin, B2 every other day, and B3 once a week with media changes. B1 confluency decreased, while B2 and B3 increased daily, except for B2 by day 9.

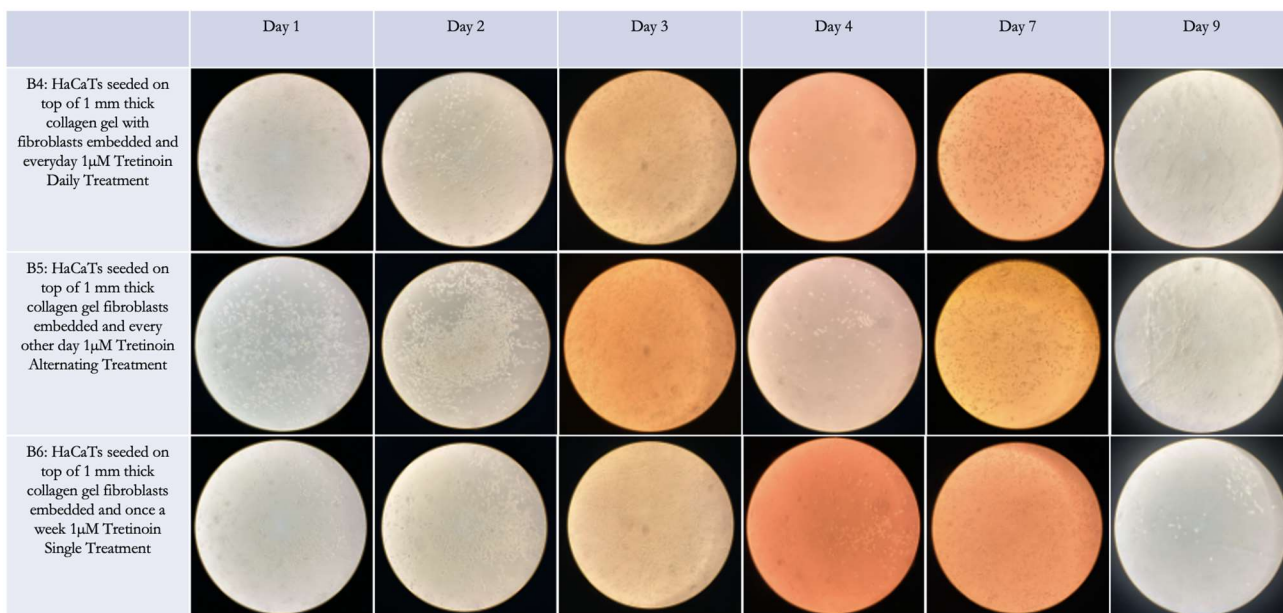


Figure 4. This figure shows daily imaging of HaCaT cells on 1 mm collagen gel with embedded fibroblasts. B4 received daily 1 µM tretinoin, B5 every other day, and B6 once a week with media changes. B4 and B6 confluency decreased, while B5 increased daily. The cells showed unhealthy morphology and detachment by the end of the experiment.

The adapalene treatments produced much healthier cells with better morphology and higher proliferation rates. The images that correlate to the adapalene treatment wells can be found in Figure 5 and Figure 6 for the younger and aged treatments respectively. Again, the wells treated less frequently with adapalene were more confluent and healthier. The confluence of younger

treatments is significantly more confluent than the older treatments, but the seeding density of the older treatments was lower than the younger treatments. This may be because there was more room for the cells to proliferate and grow since the seeding density was lower.

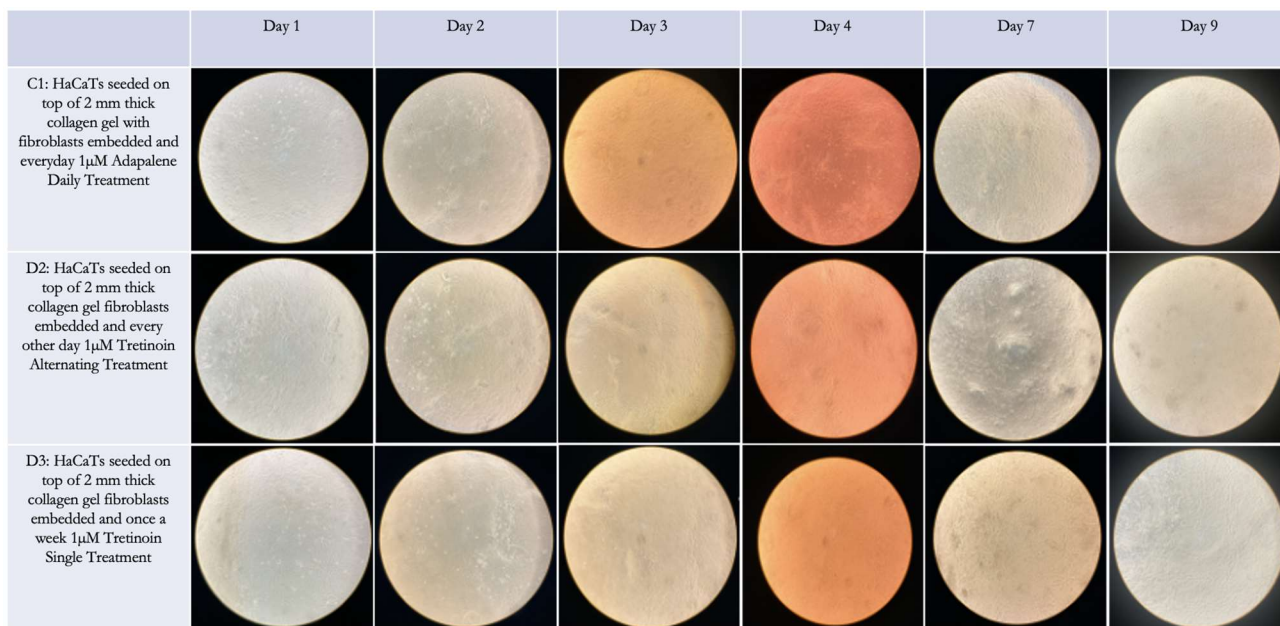


Figure 5. Daily imaging of HaCaTs on 2 mm collagen gel with embedded fibroblasts. C1 received daily 1 µM adapalene, D2 every other day, and D3 once a week with alternate-day complete media replacements. Wells remained consistently healthy and confluent throughout the experiment.

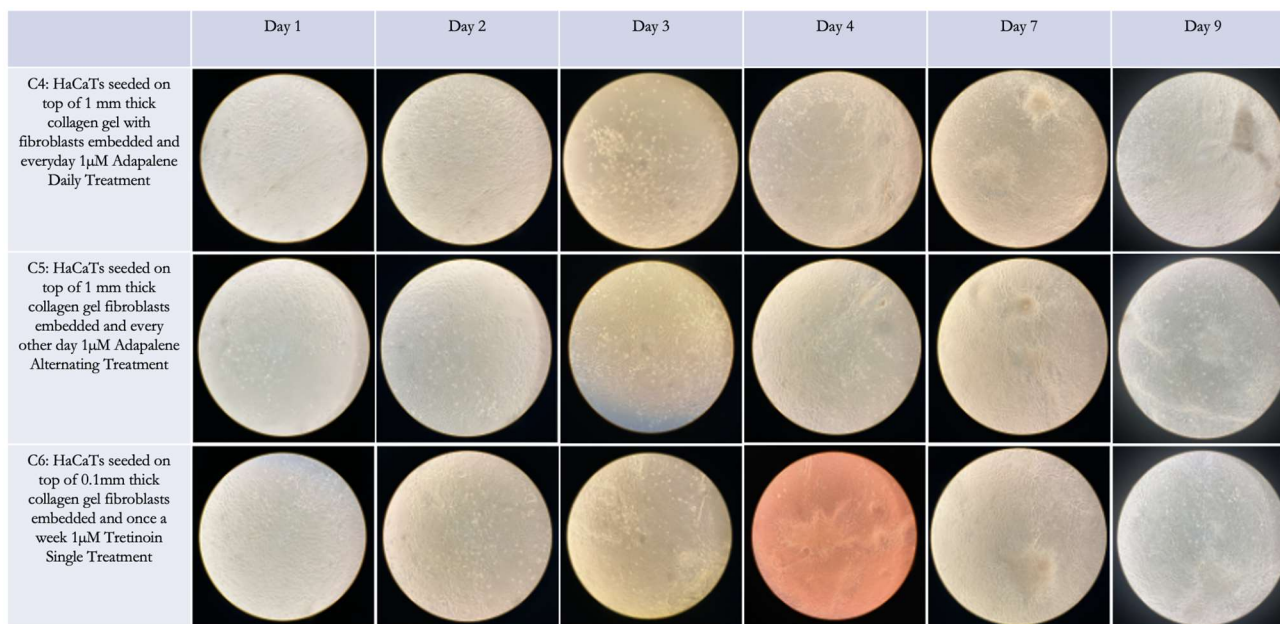


Figure 6. Daily imaging of HaCaTs on 2 mm collagen gel with embedded fibroblasts. C4 received daily 1 µM adapalene, C5 every other day, and C6 once a week with alternate-day complete media replacements. Despite lower initial cell density, all wells showed healthy and confluent cells by the experiment's end.

On the final day, the wells were imaged to assess the health of the cells, and the images can be found in **Figure 7**. Overall, the cells from the control model of HaCaTs with no retinoid treatments, were healthy and confluent. Control models with fibroblasts embedded were more confluent than those without, indicating a possible relationship between fibroblast and HaCaT proliferation. The models treated with tretinoin had cells that were less confluent than those treated with adapalene. Tretinoin is more concentrated than adapalene, so the effects are stronger. Overall, the adapalene treatment, no matter the treatment frequency or younger or aged model, had healthy and very confluent cells. For the younger adapalene treatments, as the

treatment frequency decreased, the cell confluency increased. However, for the aged adapalene treatments, as the treatment frequency decreased, the cell confluency decreased.

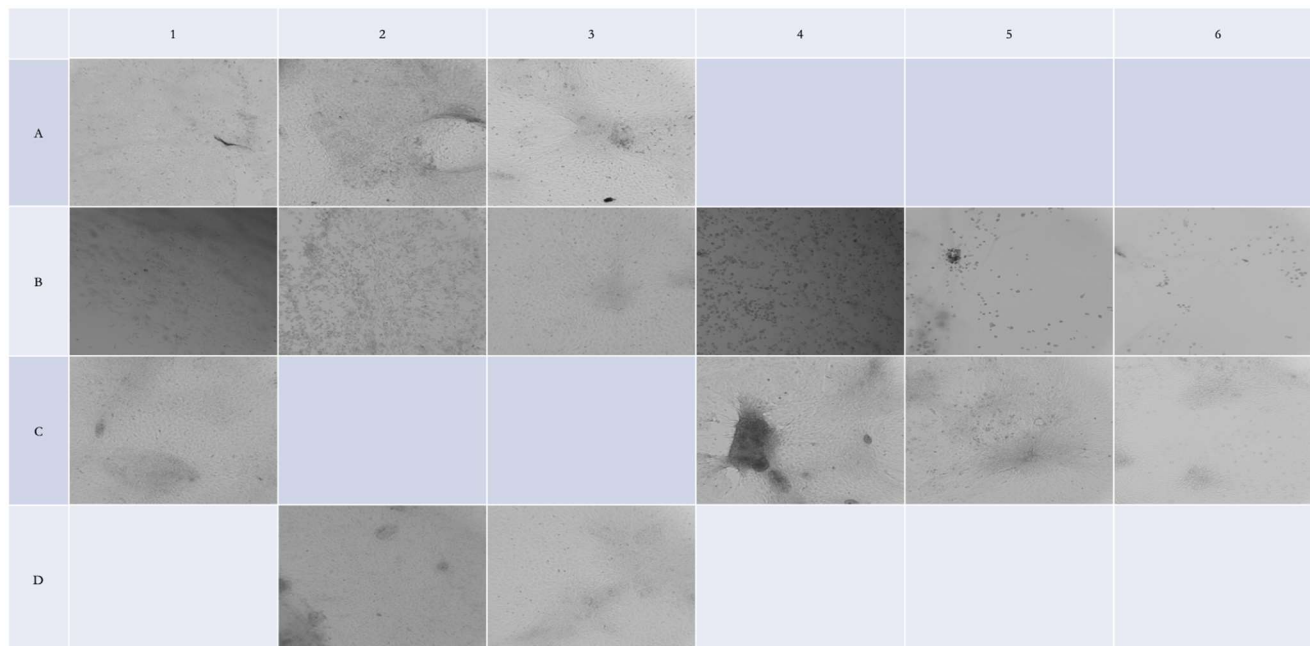


Figure 7. Final day images of HaCaTs on 3D collagen gel wells. A1, A2, and A3 show healthy and confluent cells. In the young tretinoin models (B1, B2, B3), cell density varies, with B2 showing unhealthy morphology. Aged tretinoin models (B4, B5, B6) exhibit different cell densities and morphologies, with B5 and B6 showing unhealthy cells.

Effects on Cell Proliferation

The Alamar Blue resazurin assay provided a quantitative reflection of cell metabolism for each condition, which correlates to proliferation of live cells. **Table 2** shows the contents of the Alamar Blue plate along with the associated reference-subtracted fluorescence readouts. Numbers under each description represent these reference-subtracted fluorescence values obtained for the well. 2 mm gel models used the 2 mm controls for reference subtraction, while 1 mm gel wells used the controls with 1 mm gels. Media only wells were self-subtracted. Wells are color coded according to fluorescence values, with the brightest red color representing the lowest reference-subtracted fluorescence, associated with the least number of live cells, while the brightest green color representing the highest reference-subtracted fluorescence, associated with the greatest number of live cells. A detailed key for the color code is included to the right of the table. Control wells present in both the T=0 and treatment plate included gels with and without embedded fibroblasts. The presence of fibroblasts in the collagen gel matrix correlated to greater proliferation and subsequently greater HaCaT cell proliferation. A fluorescence of 649,675 RFU was recorded with embedded fibroblasts in the 2 mm gel, and only 176,057 RFU for the same gel thickness without fibroblasts (A1).

	1	2	3	4	5	6
A	2mm CG + H + M 176057	2mm CG(F) + H + M 649675	1mm CG(F) + H + M 477158	2mm CG + M 0	1mm CG + M 0	M only 0
B	x	x	x	x	x	x
C	2mm CG + H + M 1132518	2mm CG(F) + H + M 2083877	1mm CG(F) + H + M 2085823	2mm CG + M 0	1mm CG + M 0	M only 0
D	2mm CG(F) + H + T daily 197042	2mm CG(F) + H + T alternating 740699	2mm CG(F) + H + T single 1180820	1mm CG(F) + H + T daily 98911	1mm CG(F) + H + T alternating 631870	1mm CG(F) + H + T single 1186139
E	2mm CG(F) + H + A daily 1983226	2mm CG(F) + H + A alternating 8319	2mm CG(F) + H + A single 1780495	1mm CG(F) + H + A daily 864618	1mm CG(F) + H + A alternating 2033607	1mm CG(F) + H + A single 1910548

Highest Fluorescence

Lowest Fluorescence

M = Complete DMEM
 CG = Collagen Gel
 CG(F) = Collagen Gel with NIH-3T3 fibroblast cells embedded
 H = HaCaT cells
 A = Adapalene treatment
 T = Tretinoin treatment

Table 2. Alamar Blue well contents and associated reference-subtracted fluorescence values.

Table 2 shows the well contents for the Alamar Blue assay from the T=0 control plate, in row A, and the original treatment plate, in rows C through E. Numbers under each description represent the reference-subtracted fluorescence values obtained for the well. 2 mm gel wells used the controls in wells A4 and C4 for reference subtraction, while 1 mm gel wells used the controls in wells A5 and C5. Wells are color coded according to fluorescence values, with the brightest red color representing the lowest reference-subtracted fluorescence, associated with the least number of live cells, while the brightest green color representing the highest reference-subtracted fluorescence, associated with the greatest number of live cells. A detailed key for the color code and abbreviations is included to the right of the Table.

To better visualize and compare fluorescence values for the different control and treatment conditions, data was graphed in **Figures 8 and 9**. The graphed values are referenced subtracted in the same manner as previously described. **Figure 8** depicts the primary T=0 controls, 2 mm (young) and 1 mm (old) collagen gels embedded with fibroblasts, seeded with HaCaTs on top, and DMEM media treatment. This assay was completed prior to treatment and after 3 days of seeding. The young skin model was seeded with higher HaCaT density and shows a significantly higher fluorescence, correlating with higher cell viability. Conversely the old skin model shows lower fluorescence, and less cells. These controls act as a basis for the following young and old skin models that were treated. **Figure 9** includes data for the treated models with HaCaT proliferation based on an Alamar Blue assay fluorescence readout plotted against frequency of tretinoin or adapalene treatment for young (2 mm) skin in the left panel and old (1 mm) skin in the right panel. 5 days indicates 5 consecutive days of treatment, 3 days indicates alternate days of treatment, and 1 day indicates a single first day of treatment during the course of 5 days. The well for alternate day adapalene treatment for the young skin dried causing cell death and an inaccurate readout, so this data point is not representative of the true effects of the treatment, so it was removed from **Figure 9a**; however, these data correlate with the microscopy results which show very few viable cells after the well dried. Ignoring this data point, the adapalene appears to increase proliferation with increased treatment frequency for young skin, but it does not meet or surpass the fluorescence measured for the untreated control. Adapalene with the old skin model shows alternate day treatment to have the greatest fluorescence with a value nearly equivalent to the untreated control (2,053,607 RFU and 2,085,823 RFU, respectively). Tretinoin treatment in both young and old skin shows an interesting linear trend. When a simple linear regression is made for these points, there is a nearly perfect fit in both the young and old skin, with R² values of 0.9963 and 0.9999, respectively. Tretinoin performed better when only treated once a week, but in both models this single treatment reduces cell viability by about 43% compared to the untreated control. More significant toxicity was observed if treated more than once per week.

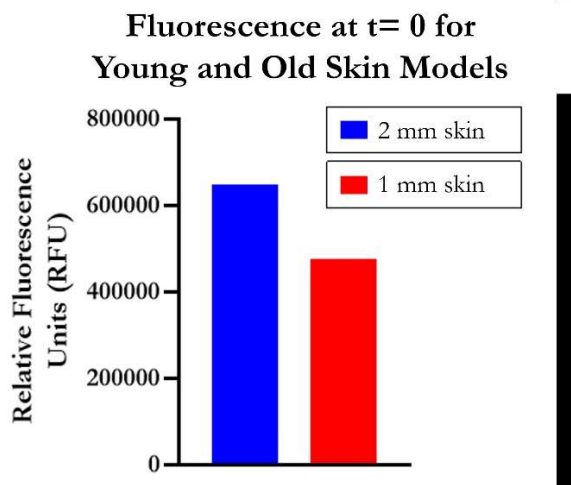


Figure 8. This Figure shows the fluorescence detected using an Alamar Blue assay for the young (2 mm) and old (1 mm) skin models at T=0, prior to treatment and 3 days after seeding. Cell metabolism was more active at the higher seeding density, as expected.

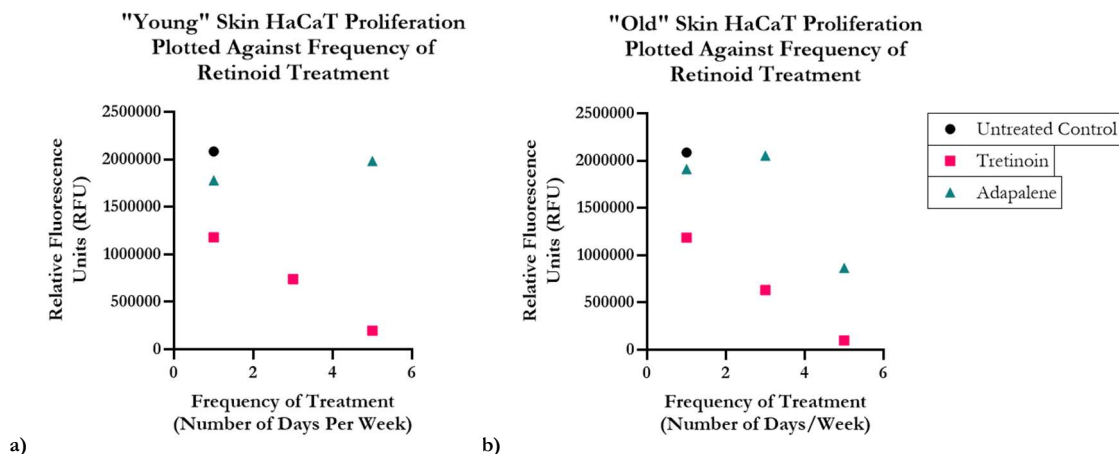


Figure 9. These Figures depict HaCaT proliferation measured by Alamar Blue assay fluorescence, plotted against tretinoin or adapalene treatment frequency for a) young (2 mm) and b) old (1 mm) skin. Note: Alternate-day adapalene data for young skin is missing due to well drying out. When a simple linear regression model is fitted for the tretinoin points on each graph, the R^2 values are 0.9963 and 0.9999, respectively.

DISCUSSION

As human skin ages, it has been shown the epidermal and dermal layers thin due to the weakening of the junction between those layers.¹⁵ It is also known that different types of retinoic acid help human skin rejuvenation and providing anti-aging effects by increasing collagen production and cell proliferation while exfoliating and removing dead skin cells. It is necessary to understand the proliferation rate provided by differing treatment frequencies with retinol products in 3D cell culture that mimics human skin.¹⁸ The data gathered for this paper shows the proliferation rates of different treatment frequencies of retinol products through an Alamar Blue proliferation assay and visual observations during and after treatment on HaCaT cells.

The results show the frequency of treatment, and the type of treatment affects the proliferation rate of the HaCaT cells. Retinoids like adapalene and tretinoin are known to increase cell turnover by promoting differentiation of skin cells, which can lead to shedding of dead skin cells and the growth of new ones.¹⁸ Adapalene and tretinoin are known to cause differentiation in HaCaT cells, differentiation of cells can lead to a slower rate of cell division and an overall decrease in cell proliferation; however, the treatment may also have induced apoptosis. When looking at the data of the untreated control wells, they have a much higher proliferation rate than those of the adapalene or tretinoin treated wells. Throughout the daily images, it can also be seen how the tretinoin treatment causes the cells to change their morphology. This may be because tretinoin is inherently more concentrated and thus more toxic to the cells than adapalene.

In addition, the side effects and safety profiles of each treatment should be considered when assessing the results. Tretinoin is known to cause skin irritation and sensitivity, while adapalene is generally better tolerated.¹⁸ The effects of retinoids on cell turnover and proliferation depend on the concentration of the treatment and duration of treatment. When the cells were treated with adapalene they had a healthier, more confluent appearance than when treated with tretinoin. Again, since tretinoin is more concentrated, it probably leads to higher rates of cell turnover, resulting in more dead cells that were lifted from the gel and were removed, without new cells being turned over in the skin cycle. The higher number of dead cells through more concentrated and more frequent treatments can be seen quantitatively in **Figure 9a** and **Figure 9b**. In this figure, tretinoin-treated models are significantly less confluent than the adapalene and control counterparts. When treated once a week and supplemented with a complete medium, more nutrients and growth factors were introduced to the cells to help the surviving cells heal and proliferate. This led to an increase in cell number over time in the once-a-week treatment.

Further studies need to be done to obtain proliferation data that incorporates HaCaT production instead of purely HaCaT growth from a set cell density. Recommendations for future experiments would be to perform a viability assay to determine the percentage of live and dead cells as well as immunostaining to assess the effects on cell differentiation between the two treatments. For future experiments, observing the healthy proliferation rate of HaCaTs and supplementing additional HaCaTs into the treatment wells to mimic cell turnover during treatment would help make the experiment and skin model more accurate. Another aspect that can be considered is the elasticity of the model since loss in elasticity is a sign of aging skin.

CONCLUSIONS

When looked at in tandem, the data from the Alamar Blue proliferation assay and the daily visual observations indicate the proliferation of HaCaTs under retinoic acid conditions is not improved. However, this is due to the set-up of the model, the

model fails to account for the loss of elasticity associated with ageing skin. It has also been shown that retinoic acid does increase the proliferation of cells in *in vivo* conditions because the body is producing more HaCaT cells underneath the layer that retinoic acid removes. When used on human skin, retinol products exfoliate and remove the top layer of skin to bring the bottom layers of skin to the surface faster to increase cell turnover. Since there is nothing in the model that allows new HaCaTs to be brought up from below the surface, proliferation could not be measured precisely. From this experiment design and preliminary data, the optimal frequency and duration of retinoid treatment cannot be concluded as there were errors in the treatment that were not consistent between each treatment well. Regardless, this study addressed the gap by creating a 3D model of human aging skin while investigating the proliferation rate of human epidermal tissue cells and the retinoic acid effects.

ACKNOWLEDGMENTS

Samantha Lopez and Olivia Atkins thank the WPI Biology and Biotechnology department, the EmpOwER grant awarded by the Women's Impact Network for funding these research initiatives, Dr. Catherine Whittington, and Dr. Jeannine Coburn (WPI Department of Biomedical Engineering) for consultation regarding materials and model construction, as well as the Cell Engineering Equipment Suite (CERES) at WPI for allowing access to fluorescence and imaging instrumentation.

REFERENCES

1. ATCC. (2022) NIH/3T3. Retrieved January 20, 2023 from <https://www.atcc.org/products/crl-1658>
2. Colombo, I., Sangiovanni, E., Maggio, R., Mattozzi, C., Zava, S., Corbett, Y., Fumagalli, M., Carlino, C., Corsetto, P. A., Scaccabarozzi, D., Calvieri, S., Gismondi, A., Taramelli, D., & Dell'Agli, M. (2017) HaCaT Cells as a Reliable In Vitro Differentiation Model to Dissect the Inflammatory/Repair Response of Human Keratinocytes. *Mediators of Inflammation*, 2017, 7435621. <https://doi.org/10.1155/2017/7435621>
3. Wilson V. G. (2014) Growth and differentiation of HaCaT keratinocytes. *Methods in Molecular Biology (Clifton, N.J.)*, 1195, 33–41. https://doi.org/10.1007/7651_2013_42
4. ATCC. (2022) Keratinocytes. Retrieved January 22, 2023, from <https://www.atcc.org/cell-products/primary-cells/keratinocytes#t=productTab&numberOfResults=24>
5. Dick, M. K., Miao, J. H., & Limaiem, F. (2022) Histology, Fibroblast. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK541065/>
6. Plikus, M. V., Wang, X., Sinha, S., Forte, E., Thompson, S. M., Herzog, E. L., Driskell, R. R., Rosenthal, N., Biernaskie, J., & Horsley, V. (2021) Fibroblasts: Origins, definitions, and functions in health and disease. *Cell*, 184(15), 3852–3872. <https://doi.org/10.1016/j.cell.2021.06.024>
7. Rhee, S. (2009) Fibroblasts in three dimensional matrices: Cell migration and matrix remodeling. *Experimental & Molecular Medicine*, 41(12), Article 12. <https://doi.org/10.3858/emm.2009.41.12.096>
8. Chen, H., Xue, L., Gong, G. et al. (2022) Collagen-based materials in reproductive medicine and engineered reproductive tissues. *Journal of Leather Science and Engineering* 4, 3. <https://doi.org/10.1186/s42825-021-00075-y>
9. Caliarì, S., Burdick, J. (2016) A practical guide to hydrogels for cell culture. *Nature Methods* 13, 405–414. <https://doi.org/10.1038/nmeth.3839>
10. Alamán-Díez, P., García-Gareta, E., Arruebo, M., Pérez, María Ángeles. (2023) A bone-on-a-chip collagen hydrogel-based model using pre-differentiated adipose-derived stem cells for personalized bone tissue engineering. *Journal of Biomedical Materials Research*. 2023; 111(1): 88- 105. doi:10.1002/jbm.a.37448
11. Andersen, T., Auk-Emblem, P., & Dornish, M. (2015) 3D Cell Culture in Alginate Hydrogels. *Microarrays*, 4(2), 133–161. <https://doi.org/10.3390/microarrays4020133>
12. Yao, L., Tran, K., & Nguyen, D. (2022) Collagen Matrices Mediate Glioma Cell Migration Induced by an Electrical Signal. *Gels*, 8(9), 545. <https://doi.org/10.3390/gels8090545>
13. Chen, Q., & Ross, A. C. (2004) Retinoic acid regulates cell cycle progression and cell differentiation in human monocytic THP-1 cells. *Experimental Cell Research*, 297(1), 68–81. <https://doi.org/10.1016/j.yexcr.2004.02.017>
14. Breitkreutz, D., Stark, H. J., Plein, P., Baur, M., & Fusenig, N. E. (1993) Differential modulation of epidermal keratinization in immortalized (HaCaT) and tumorigenic human skin keratinocytes (HaCaT-ras) by retinoic acid and extracellular Ca²⁺. *Differentiation*, 54(3), 201–217. <https://doi.org/10.1111/j.1432-0436.1993.tb01602.x>
15. Yokoshiki, S., Maeda, M., & Saijo Y., (2017) High resolution facial skin imaging with three-dimensional ultrasound microscope. *Proceedings of Meetings of Acoustics*. 32, 020015 <https://doi.org/10.1121/2.000>
16. Dos Santos, M., Metral, E., Boher, A., Rouselle, P., Thepot, A., Damour, O. (2015) In vitro 3-D model based on extending time of culture for studying chronological epidermis aging, *Matrix Biology*, 47, 85–97. <https://doi.org/10.1016/j.matbio.2015.03.009>.
17. Weinmüller, R., Zbiral, B., Becirovic, A., Stelzer, E. M., Nagelreiter, F., Schosserer, M., Lämmermann, I., Liendl, L., Lang, M., Terlecki-Zaniewicz, L., Andriotis, O., Mildner, M., Golabi, B., Waidhofer-Söllner, P., Schedle, K., Emsenhuber, G., Thurner, P. J., Tschachler, E., Gruber, F., & Grillari, J. (2020) Organotypic human skin culture models constructed with

- senescent fibroblasts show hallmarks of skin aging. *Npj Aging and Mechanisms of Disease*, 6(1), Article 1. <https://doi.org/10.1038/s41514-020-0042-x>
18. Cleveland Clinic medical professional. (2022, June 17) *Retinol: Cream, serum, what it is, benefits, how to use*. Cleveland Clinic. Retrieved February 6, 2023, from <https://my.clevelandclinic.org/health/treatments/23293-retinol>
 19. Tolaymat L., Dearborn H., Zito PM. (2022) Adapalene. *StatPearls*.
 20. Yoham, A., Casadesus D., (2022) Tretinoin. *StatPearls*.
 21. Shao, Y., He, T., Fisher, G. J., Voorhees, J. J., & Quan, T. (2017) Molecular basis of retinol anti-ageing properties in naturally aged human skin in vivo. *International Journal of Cosmetic Science*, 39(1), 56–65. <https://doi.org/10.1111/ics.12348>
 22. Babamiri, K., & Nassab, R. (2010) Cosmeceuticals: The Evidence Behind the Retinoids. *Aesthetic Surgery Journal*, 30(1), 74–77. <https://doi.org/10.1177/1090820X09360704>
 23. Fujisaki, H., Futaki, S., Yamada, M., Sekiguchi, K., Hayashi, T., Ikejima, T., & Hattori, S. (2018) Respective optimal calcium concentrations for proliferation on type I collagen fibrils in two keratinocyte line cells, HaCaT and FEPE1L-8. *Regenerative Therapy*, 8, 73–79. <https://doi.org/10.1016/j.reth.2018.04.001>
 24. Farage, M. A., Miller, K. W., Elsner, P., & Maibach, H. I. (2013) Characteristics of the Aging Skin. *Advances in Wound Care*, 2(1), 5. <https://doi.org/10.1089/wound.2011.0356>

ABOUT STUDENT AUTHORS

Samantha Lopez and Olivia Atkins graduated in May 2023 with a Bachelor of Science in Biomedical Engineering and Biology and Biotechnology, respectively. They both plan on furthering their education and pursuing a master's degree in their respective fields.

PRESS SUMMARY

Human skin aging is characterized by thinning, loss of elasticity, and wrinkles. Keratinocytes, the most common type of skin cell, and fibroblasts, cells present in the stroma beneath the skin's surface, both play a role in aging. Using these cell lines in research can reveal a deeper understanding of skin function and changes in aging. 3D cell culture techniques provide an opportunity to use these cell lines in a model that can more accurately mimic human skin compared to 2D models. Treatment of aging skin is of interest to both medical and consumer communities. Retinoic acid (RA) is a derivative of both vitamin A and retinol that assists in cell growth and immune functions. Over the counter (OTC) and prescription retinoids are common topical products used for anti-aging and acne treatments. This study seeks to determine the impact of topical retinoid creams on skin cell growth and characteristics in 3D cell culture models mimicking aged and unaged human skin. NIH-3T3 fibroblasts were embedded in a 3D collagen matrix of varying thickness, and HaCaT keratinocytes were seeded on top of the matrix at varying seeding densities. Equivalent concentrations of 0.025% tretinoin and 0.1% adapalene topical creams were prepared in cell medium and used to treat cells daily, on alternate days, or just once during a week-long period. Alamar Blue proliferation assays and microscopy provided quantitative and qualitative data regarding the effects of the retinoid products on the skin models. Tretinoin treatment killed the cells, and adapalene treatment, while resulting in significantly more live cells than tretinoin, did not exceed the proliferation of the untreated control. It is understood that retinol increases cell turnover by killing cells rapidly, so it is proposed that in this model, the rate of proliferation does not overcome the rate of cell death. Considering the data is only preliminary, future studies should focus on creating a 3D model that can more accurately represent the skin environment in the human body where cells can be readily grown.

Increasing Access to a Four-Year College: Impacts of a California State University Guaranteed Admission Program on College Enrollment Rates

Joleen Chiu*

Department of Economics, University of California, Los Angeles, Los Angeles, CA

<https://doi.org/10.33697/ajur.2023.094>

*Student: joleENCHIU@ucla.edu**

Mentor: j.londonovelez@econ.ucla.edu

ABSTRACT

Guaranteed admission programs are a type of college access program that provide students who meet certain criteria (e.g., a minimum GPA) with guaranteed admission to one or more colleges. This paper studies guaranteed admission agreements between California State University San Marcos (CSUSM) and its local school districts to evaluate if smaller-scale, local guaranteed admission programs have comparable impacts on college enrollment rates to previously studied state-wide programs. Employing a regression discontinuity (RD) design around the program's GPA cutoff conditional on students satisfying other program requirements, this paper finds that the program significantly increased enrollment at CSUSM and at any California State University (CSU) and increased (not significantly) the likelihood of students enrolling at four-year institutions compared to two-year alternatives. In addition, the program disproportionately affected students from underrepresented backgrounds (e.g., first-generation, nonwhite, and low-income backgrounds), suggesting that local guaranteed admission programs also have the potential to increase representation at four-year public institutions and encourage underrepresented students to enroll at higher-quality postsecondary institutions.

KEYWORDS

Guaranteed Admission; College Enrollment; College Admissions; College Access; Higher Education; California State University; Education Policy; Local Policy

INTRODUCTION

Obtaining a college degree typically opens the door to numerous benefits, including a wider selection of occupational choices and more economic stability. Aside from being able to access over 160 different types of occupations that require a bachelor's degree,¹ American college graduates experience, on average, lower unemployment rates and higher hourly wages compared to Americans whose highest level of educational attainment is a high school diploma.²⁻⁴ Yet the benefits of having a college degree are distributed unequally in the United States. Students who identify as an underrepresented racial minority, come from a low-income background, or whose parents do not have a college degree are significantly less likely to pursue and obtain a college degree.⁴⁻⁶

Consequently, policymakers have experimented with different approaches to increase college enrollment, especially among underserved community members, such as implementing financial aid and guaranteed admission programs. Financial aid policies, such as the national need-based Pell Grant and state-based programs like the Cal Grant and Maryland's Educational Assistance Grant, directly address issues surrounding the cost of college and commonly target low-income students.⁷ On the other hand, guaranteed admission policies grew in popularity as multiple states in the U.S. banned the use of affirmative action (the consideration of race and ethnicity in college admissions). States like Texas

and California began providing students who graduate near the top of their high school class with guaranteed admission to in-state public universities after the ban of affirmative action in their state as an attempt to maintain diversity in public higher education.^{8,9} As of 2021, 13 states have implemented a state-wide guaranteed admission program.¹⁰

This paper focuses on the second type of policy: guaranteed admission programs. Existing research finds that the Texas and California guaranteed admission programs significantly increased enrollment and graduation rates at high-quality, in-state public universities among historically underrepresented students in higher education, such as those from lower-income backgrounds.^{8,11} In contrast, a guaranteed admission program in Idaho involving a mixture of two-year and four-year institutions significantly increased two-year college enrollment rates but did not significantly impact the enrollment rates of low-income students.¹²

These previous findings have important implications for addressing inequality as the policies in Texas and California primarily benefitted underrepresented students following the end of affirmative action in those states. However, little is known about how the effects of a locally targeted guaranteed admission program may differ from these previously studied state-wide policies. For instance, some institutions like the California State University (CSU) system studied in this paper have a stronger focus on serving its local communities (*e.g.*, by providing priority admissions to local school districts), potentially making the guaranteed admission option more attractive to local students.¹³

Furthermore, previous papers do not consider how variations in guaranteed admission programs' requirements may result in differing enrollment outcomes. The previous findings mentioned above suggest that policies targeting high-achieving minority students may be more effective in encouraging them to enroll at higher-quality institutions (as seen in the Texas and California policies) whereas policies that expand their scope to include more students may be more effective in increasing the overall college enrollment rate (as seen in the Idaho policy).^{8,11,12} Additional research can help establish this trend and extend it to local guaranteed admission programs.

This paper provides insight into the effectiveness of local guaranteed admission programs and how program requirements may shape enrollment outcomes by studying a guaranteed admission program between California State University San Marcos (CSUSM) and its surrounding school districts. CSUSM signed Memorandums of Understanding (MOUs) with 10 local school districts between 2009 to 2015,^A which guarantee students who graduate from a partner school district admission to the university if they satisfy the standard CSU admission requirements (*i.e.*, complete the A-G courses and pass the entry-level math and English exams), graduate with a minimum Grade Point Average (GPA) of 3.0, and take the SAT Reasoning Test (SAT I) or ACT exam. This CSUSM program differs from previously studied guaranteed admission programs since it focuses on encouraging students from a small, local community to enroll at a certain college compared to state-wide programs that usually involve multiple public universities. Additionally, compared to other four-year institutions, CSU schools like CSUSM serve a larger proportion of minority and low-income students, potentially making them more effective at targeting underrepresented students through the MOUs.¹⁴ Moreover, during the years studied in this paper (2017 to 2019), CSUSM did not require students to provide personal statements or background on their extracurricular involvement in their application, simplifying the application process and possibly encouraging more students to take advantage of the guaranteed admission program.¹⁵⁻¹⁷

The author collects novel data directly from two school districts that have signed MOUs with CSUSM. The data consists of 21,191 graduates between 2016 to 2022 and describes their high school performance (*e.g.*, GPA, SAT/ACT completion status, and A-G course completion status), demographic information (*e.g.*, race, gender, and parents' highest level of education attainment), and the first postsecondary institution in the U.S. enrolled at after high school (if any). This information is combined with data from the Integrated Postsecondary Education Data System (IPEDS) to classify each postsecondary as a four-year or two-year institution. The analysis sample is restricted to 4,502^B students who graduated between 2017 to 2019 and satisfied the CSU A-G course requirement and the SAT I/ACT requirement for guaranteed admission.

Using a regression discontinuity (RD) analysis around the 3.0 GPA cutoff for the CSUSM MOUs, this paper finds that the CSUSM MOUs do not affect overall enrollment rates at any postsecondary. However, students just above the GPA cutoff are significantly more likely to enroll at CSUSM or any CSU in lieu of two-year institutions. Furthermore, these results are driven mostly by students whose parents did not attain a college degree, nonwhite students, and low-income students, closing the college enrollment gaps between students whose parents did/did not attain a college degree and the CSUSM and CSU enrollment gaps by income. These results indicate that the CSUSM MOUs are encouraging underrepresented students to shift away from enrolling at two-year institutions towards enrolling at four-year institutions, particularly towards the CSUs. Hence, the CSUSM MOUs are increasing underrepresented student enrollment at higher-quality institutions and closing historic gaps in enrollment in the CSU system. These results hold even when including the graduating class of 2020 in the analysis sample or when doubling the RD bandwidth sizes. The author also validates these findings using a placebo sample of 1,882 graduates from 2016 and finds no significant increases in any postsecondary, any four-year, CSUSM, or CSU enrollment and no significant decreases in two-year enrollment for this sample above the GPA cutoff.

This paper contributes new findings on the effectiveness of guaranteed admission programs and presents insight into how the effectiveness of local programs may differ from state-wide programs. The CSUSM MOUs target a much smaller community (10 school districts) compared to state-wide programs. Therefore, the CSUSM MOUs may be especially appealing to students considering a local two-year alternative or could potentially incentivize students on the brink of attending college to enroll in a local, guaranteed four-year option. Since CSUSM shares a common application with other CSU campuses,¹⁶ this paper provides insight into how the MOUs may expand enrollment at other CSUs by incentivizing students to complete the CSU application process. In addition, due to local guaranteed admission programs being much smaller-scale compared to state-wide programs, they may be less likely to displace other qualified students.

This paper also contributes more broadly to the literature studying inequality in higher education access and enrollment. The restrictiveness of the conditions for being eligible for the CSUSM MOUs is situated between the top percent policies (*i.e.*, the Texas and California policies) and the Idaho direct admissions program. The CSUSM MOUs do not restrict access to the most qualified students like in the top percent policies but still require students to meet a 3.0 GPA requirement. Hence, this paper could provide further insight into how a program's requirements accessibility could impact the effectiveness of the program or the group(s) of students the program targets.

Prior Literature on Guaranteed Admission Programs

Guaranteed admission programs rose in popularity with the ban on affirmative action in several states. As a result, guaranteed admission programs can be an important facilitator in increasing the college enrollment rates of historically underrepresented students and their social mobility. On the other hand, existing concerns about affirmative action programs "displacing" other qualified students or leading to a "mismatch" of underrepresented students at higher-quality institutions compared to the institutions they would have enrolled at otherwise carry over to guaranteed admissions.^{8, 11, 18} Therefore, previous literature on guaranteed admission programs studies not only the programs' effects on college enrollment, graduation rates, and future earnings but also which demographics in particular are benefiting, or potentially harmed, the most under such policies.

The most studied guaranteed admission program is the Texas Top Ten Percent Program (TTP). After the University of Texas's consideration of race in college admission decisions was deemed unconstitutional in 1996, Texas lawmakers introduced the TTP as an attempt to maintain diversity in higher education, providing guaranteed admission to any Texas public university to Texan students who rank in the top 10 percent of their high school class beginning with the class of 1998.^{8, C} Since the inception of the TTP, students who graduate from high schools with historically low representation in the University of Texas (UT) schools, such as high schools with large proportions of students who live under the poverty line, receive free or reduced lunch, or identify as an underrepresented minority, have seen an increase in enrollment in the UT system.¹⁹ Past research also finds that students respond to the TTP as public universities like the

University of Texas at Austin (UT Austin) and Texas A&M University saw an increase in applications while competing private institutions like Rice University and Southern Methodist University saw a decline in applications after the implementation of the TTP.²⁰

Black *et al.* (2023) expand upon these previous findings by exploring the “displacement” and “mismatch” questions using the Texas public flagship university, UT Austin.⁸ They find that not only were underrepresented students more likely to enroll at UT Austin, but they were also more likely to graduate from UT Austin and earn higher earnings under the TTP. This provides evidence against concerns of a potential “mismatch” of underrepresented minorities at high-quality institutions and suggests that the TTP has positive impacts on income mobility. Although they did find that this policy displaced students who attended high schools with historically high enrollment rates at UT Austin but did not qualify for the TTP, the displaced students’ eventual college graduation rates and post-graduate earnings were not negatively impacted.

California introduced a similar policy, Eligibility in the Local Context (ELC), in 2001, guaranteeing students who graduate in the top four percent of a California high school admission to at least one University of California (UC) institution.^{11,D} Bleemer (2021) found that the ELC affected enrollment at four UC campuses in particular, henceforth referred to as the “Absorbing” UCs: Davis, Irvine, Santa Barbara, and San Diego.¹¹ Students who graduated from the bottom half of Californian high schools by SAT score were more likely to enroll at and graduate from Absorbing UCs and less likely to enroll at less selective UCs, a CSU, or a California community college. These students were also more likely to identify as Black or Hispanic and more likely to come from families with below-median incomes, providing further evidence against the “mismatch” theory.

While the TTP and ELC target high-achieving students through their minimum class rank requirement, Idaho established a direct admissions system in 2015 that determined admissions using a student’s ACT or SAT completion status and GPA instead.¹² All Idaho high school graduates were either guaranteed admission to a group of six colleges or a group of eight colleges (both of which consisted of a mixture of two-year and four-year institutions) depending on their ACT/SAT completion status, number of high school credits, and GPA.¹² As a result of this policy, college enrollment rates increased significantly for first-time in-state undergraduates, concentrating primarily among two-year, open-access institutions with no significant effects on the enrollment of Pell-eligible students.¹² This finding presents a new perspective where guaranteed admission policies that expand their focus beyond high-achieving students may be more effective in encouraging students to consider pursuing a guaranteed postsecondary education (*i.e.*, a guaranteed two-year institution in this case) and thus increasing the overall college enrollment rate, whereas top percent policies like the TTP and ELC may be more effective in improving the type of institution high-achieving underrepresented students enroll at.

The CSU System and the CSUSM MOUs

CSUSM is one of 23 campuses in the larger California State University (CSU) system, one of two public university systems in California (the other being the University of California). As of Fall 2020, CSUSM had an enrollment of 16,367 students, representing 3.27 percent of the total CSU enrollment.^E Among them, 15,040 (91.9 percent) were undergraduate students, 10,145 (62 percent) identified as female, and 9,162 (56 percent) identified as an underrepresented minority,^F which all occurred in higher proportions than the overall CSU population. Remarkably, the CSU system as a whole has a higher enrollment of underrepresented and low-income students compared to other four-year colleges.¹⁴

To apply to the CSU and the UC systems, students must complete 15 yearlong high school courses, known as the A-G courses, with a letter grade of “C” or above.^{15, 16, 21} The CSU campuses share one application, in which admission is determined by students’ high school courses and grades and SAT or ACT exam scores.^{15, 16,G} Notably, the CSUs do not require any personal statements or essays or inquire about extracurricular involvement in their application,^{15-17,H} streamlining the application process.

Between the two California public university systems, the CSU system has a stronger focus on serving its local communities by providing priority admission to their surrounding school districts and counties.¹³ In particular, CSUSM's local admission area covers 23 high school districts in the counties of Orange, Riverside, and San Diego in Southern California.^{23,1} Students from these local area school districts have priority placement in CSUSM's impacted majors (which change annually) if they meet the general CSUSM admission requirements.²⁴

In addition to providing priority admission to its local school districts, as of 2022, CSUSM has also signed Memorandums of Understanding with 10 school districts in its local admission area, known as their partner districts or "The Alliance".²⁵ Starting between the classes of 2009 to 2019, high school students who graduate from a CSUSM partner school district are guaranteed admission to CSUSM if they^J

- complete the A-G courses with a grade of "C" or higher;
- achieve a passing score on the English Placement Test (EPT) and Entry Level Math (ELM);^K
- take the SAT Reasoning Test (SAT I) or ACT exam (no minimum score required); and
- graduate with a minimum cumulative GPA of 3.0.^{27-36,L}

Table 1 compares the 10 school districts that have signed MOUs with CSUSM. Notably, the school districts in the Alliance require students to complete most of the A-G courses as a graduation requirement,³⁷⁻⁴⁶ assisting students with meeting the CSUSM MOU requirements. This paper focuses on two of the 10 school districts where the author could obtain data: Murrieta Valley Unified School District (MVUSD) and Vista Unified School District (VUSD).

METHODS AND PROCEDURES

Data

The author observes individual-level student data obtained directly from the two districts studied in this paper, which consist of 14,246 graduates between 2016 to 2022 from Murrieta Valley Unified School District (MVUSD) and 6,945 graduates between 2018^M to 2022 from Vista Unified School District (VUSD). The two districts used a combination of data from the National Student Clearinghouse (NSC) to obtain college enrollment information along with the California Longitudinal Pupil Achievement Data System (CALPADS) and district records of student surveys and performance to obtain student high school performance and demographic information. The classes of 2020 to 2022 are omitted from the analysis due to extremely low SAT/ACT completion rates during the COVID-19 pandemic.^N In addition, the uncertainty surrounding online learning may have affected students' decisions to pursue higher education. The MVUSD class of 2016 is also omitted from the main analysis since they graduated before the CSUSM MOUs went into effect for that district.

The restricted data (to classes of 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD) contains information on 9,032 (6,112 from MVUSD and 2,920 from VUSD) students' high school performance (cumulative GPA at the time of graduation and standardized tests taken), the first postsecondary institution in the U.S. enrolled at after high school (if any), and demographic information (race, gender, and parents' highest level of educational attainment) (See **Table 2**). The data also includes free and reduced lunch status for MVUSD and whether a student is considered socioeconomically disadvantaged for VUSD. These two variables are used to create a general low-income variable for students who are on free and reduced lunch status or are considered socioeconomically disadvantaged. Lastly, this data is combined with classifications of each postsecondary institution as a two-year or four-year institution from the Integrated Postsecondary Education Data System (IPEDS).

A-G Course Requirements Included in District Graduation Requirements

School District	County	First Graduating Class Impacted by MOU	A-G Course Requirements						
			A. History (2 years)	B. English (4 years)	C. Math (3 years)	D. Science (2 years)	E. Language other than English (2 years)	F. Visual and performing arts (1 year)	G. College-preparatory elective (1 year)
Catsbad Unified	San Diego	2017	✓	✓	✓	✓	✓	✓	✓
Escondido Union	San Diego	2011	✓	✓	✓	✓			✓
Fallbrook Union High	San Diego	2015	✓	✓	✓	✓			✓
Lake Elsinore Unified	Riverside	2019	✓	✓	✓	✓			✓
Murrieta Valley Unified	Riverside	2017	✓	✓	✓	✓		✓	✓
Oceanside Unified	San Diego	2017	✓	✓	✓	✓	✓	✓	✓
San Marcos Unified	San Diego	2009	✓	✓	✓	✓	✓	✓	✓
Temecula Valley Unified	Riverside	2014-15	✓	✓	✓	✓	✓	✓	✓
Valley Center-Pauma Unified	San Diego	2011	✓	✓	✓	✓	✓	✓	✓
Vista Unified	San Diego	2017	✓	✓	✓	✓	✓	✓	✓

Table 1. Comparisons of Districts in the CSUSM Alliance. This table lists the ten school districts who have signed a guaranteed admission MOU with CSUSM along with the county they are located in, the first graduating class eligible for the MOU, and the A-G courses that students satisfy in order to graduate from each district. The graduation requirements reflect current graduation requirements. Some requirements may have changed since the introduction of the MOUs. Some districts require students to take one year of any fine art, which could satisfy either one year of visual/performing arts or one year of world language. The author considers these situations as not required to satisfy both the language other than English requirement and the visual and performing arts requirement. The paper uses a binary variable of whether or not a student satisfied the A-G courses given directly by VUSD to determine a student's eligibility for the CSUSM MOUs.³⁷⁻⁴⁶

Notably, the dataset does not include whether students completed the A-G course requirements for MVUSD students. However, MVUSD's graduation requirements are only a one-year-long course difference away from the A-G course requirements,⁴² and the majority of MVUSD graduates (65.1 percent for the class of 2020) satisfy the A-G requirement.⁴⁷ In addition, students who complete the A-G requirements are likely more interested in attending a four-year institution, especially a UC or CSU institution, suggesting that the analysis using all MVUSD students rather than the sample of MVUSD students who completed the A-G requirements is likely an underestimate of the true effect of the CSUSM MOUs.⁹

The dataset also does not include whether students obtained a passing score on the EPT and ELM exams or satisfied one of the exemptions for both school districts. Students can fulfill the EPT and ELM requirements through many other means, including but not limited to

- obtaining a minimum cumulative GPA of 3.7;
- obtaining a minimum cumulative GPA of 3.0 and enrolling in a year-long senior English class (for EPT);
- obtaining a minimum cumulative math GPA of 3.0 and enrolling in a year-long senior math class (for ELM); or
- achieving a sufficiently high SAT I critical reading score, ACT English score, Advanced Placement (AP) English (Language and Composition or Literature and Composition), AP Statistics, or AP Calculus (AB or BC) exam scores.²⁶

Students also have the opportunity to complete additional coursework to prepare them for the curriculum at CSUSM.²⁶ Therefore, this paper assumes most students are exempt from the EPT/ELM requirement or have the opportunity to fulfill the requirement through additional coursework.

Summary Statistics of Full and RD Samples

Table 2 describes summary statistics for the 9,032 high school graduates from MVUSD and VUSD between 2017 to 2019. The first column presents the summary statistics for the full sample of high school graduates from the two districts, and columns two and three present the values broken down by district. 71 percent of all graduates enrolled at a postsecondary institution, 34 percent enrolled at a four-year college, and 6.5 percent enrolled at CSUSM after high school. 50 percent of all graduates are female, 41 percent identify as Hispanic, 39 percent identify as white, nine percent identify as Asian or Pacific Islander, five percent identify as Black, and less than one percent identify as American Indian or Alaskan Native. 48 percent of all graduates are first-generation college students^P and 38 percent come from a low-income background. VUSD has a higher proportion of Hispanic students (57 percent), first-generation college students (55 percent), and low-income students (63 percent) compared to MVUSD (33, 45, and 26 percent respectively), and a lower percentage of students who enrolled at any postsecondary institution (62 percent) or any four-year college (27 percent) compared to MVUSD (75 and 38 percent respectively).

Column four of **Table 2** displays the summary statistics for the RD Sample, and columns five and six present the summary statistics for the RD sample broken down by district. In this paper, the RD sample refers to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who satisfied the A-G course requirements (VUSD only), took the SAT I or ACT exam, and graduated with a GPA between 2.73 to 3.43, which is the data-driven selected bandwidth (described in the next section). The overall RD sample is less likely to be first-generation or low-income, but otherwise demographically very similar to the overall sample of students. Because the sample of students near the 3.0 GPA is positively selected in their ability and in their interest in pursuing higher education (demonstrated through completing the SAT I or ACT exam) relative to the average student, the RD sample also has much higher college enrollment rates compared to the general sample of high school graduates from MVUSD and VUSD, with 88 percent enrolling at any postsecondary institution, 52 percent enrolling at a four-year institution, and 18 percent enrolling at CSUSM after high school. Similar to the overall sample, MVUSD students in the RD sample have higher enrollment rates at two-year and four-year colleges compared to VUSD.

	All Graduates			RD Sample		
	Overall	MVUSD	VUSD	Overall	MVUSD	VUSD
	(N = 9032)	(N = 6112)	(N = 2920)	(N = 1616)	(N = 1325)	(N = 291)
	(1)	(2)	(3)	(4)	(5)	(6)
Demographic Variables						
% Female	50.32	48.51	54.11	51.42	50.34	56.36
% American Indian or Alaskan Native	0.29	0.29	0.27	0.25	0.30	0.00
% Asian or Pacific Islander	9.23	10.31	6.99	9.96	10.57	7.22
% Black	5.12	6.32	2.60	7.05	7.77	3.78
% Hispanic	40.77	33.12	56.78	37.31	33.81	53.26
% White	39.19	43.82	29.52	38.68	40.00	32.65
% First-Generation College Students	48.33	45.12	55.03	45.30	44.00	51.20
% Low-Income	38.04	25.93	63.39	30.20	24.83	54.64
High School Performance						
Average GPA	3.03	3.08	2.92	3.13	3.13	3.15
% Took SAT Reasoning Test or ACT	53.41	56.22	47.53	100.00	100.00	100.00
% Satisfied A-G Requirement	NA	NA	44.35	NA	NA	100.00
% Satisfied EPT/ELM Requirement	NA	NA	NA	NA	NA	NA
% Eligible for Guaranteed Admission	40.00	43.49	32.71	72.59	72.38	73.54
College Enrollment Rates						
% Any Postsecondary	71.08	75.44	61.95	88.12	90.34	78.01
% Any Four-year	34.32	37.75	27.16	52.10	53.74	44.67
% Any Two-year	36.76	37.70	34.79	36.01	36.60	33.33
% CSUSM	6.53	6.54	6.51	18.07	18.04	18.21
% Any CSU	12.75	13.47	11.27	28.40	28.15	29.55
% Any UC	6.71	6.92	6.27	1.42	1.36	1.72

Table 2. Descriptive Statistics of Alliance Students. This table presents the summary statistics for the full sample of MVUSD graduates from 2017 to 2019 and VUSD graduates from 2018 to 2019 along with the summary statistics for the RD sample only. The RD sample consists of students who satisfied the A-G requirement (VUSD only), took the SAT I or ACT exam, and have a GPA between 2.73 to 3.43.

Regression Discontinuity Design

This paper evaluates the effect of the CSUSM MOUs on postsecondary enrollment rates using a regression discontinuity design around the 3.0 GPA cutoff conditioned on students who completed the A-G course requirements (VUSD only) and the SAT I or ACT exam. The effect is estimated through β_1 from a linear regression model:

$$y_i = \beta_0 + \beta_1 \cdot \mathbb{1}(GPA_i \geq 3) + \beta_2 \cdot GPA_i + \epsilon_i \tag{Equation 1.}$$

where y_i represents the outcome variable of interest, $\mathbb{1}(GPA_i \geq 3)$ is a binary variable that returns true if a student has a GPA above the cutoff of 3.0, and GPA_i is the student’s GPA.

This paper focuses on five outcome measures of college enrollment rates between different types of institutions: any postsecondary, four-year institutions only, two-year institutions only, CSUSM only, and any CSU. Although the CSUSM MOUs only directly impact admission to CSUSM, the author considers if the program increased the proportion of students who pursue a postsecondary in general and if it changes the type of institution students enroll at (*i.e.*, a shift from two-year institutions to four-year institutions). In addition, since the CSUs all share one application, this paper evaluates if the program may have encouraged students to apply to and enroll at other CSU campuses. The author then explores heterogeneity in the enrollment outcomes by parents’ education level, gender, race, and income status using separate estimates of **Equation 1** for each group of students.

The analysis follows the data-driven bandwidth selection strategy described in Calonico *et al.* (2014),⁴⁸ which recommends a bandwidth (BW) ranging from 2.73 to 3.43 when the outcome variable is any postsecondary. Therefore, the analysis focuses on the sample of graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who satisfied

the A-G course requirements (VUSD only), took the SAT I or ACT exam, and graduated with a GPA between 2.73 to 3.43 (the “RD sample”).

Validity of the RD Model

A key assumption of the RD model is that students cannot manipulate their grades to be just above the GPA cutoff. This assumption is checked in **Figure 1**, which plots the distribution of student GPAs for those who are eligible for the CSUSM MOUs (completed the A-G requirement for VUSD and completed the SAT/ACT requirement for both districts) compared to those who are ineligible. Following the McCrary (2008) test of distributional discontinuity,⁴⁹ there are no significant changes ($p=0.18$) in the density of eligible students just above the 3.0 GPA cutoff, validating the RD model.

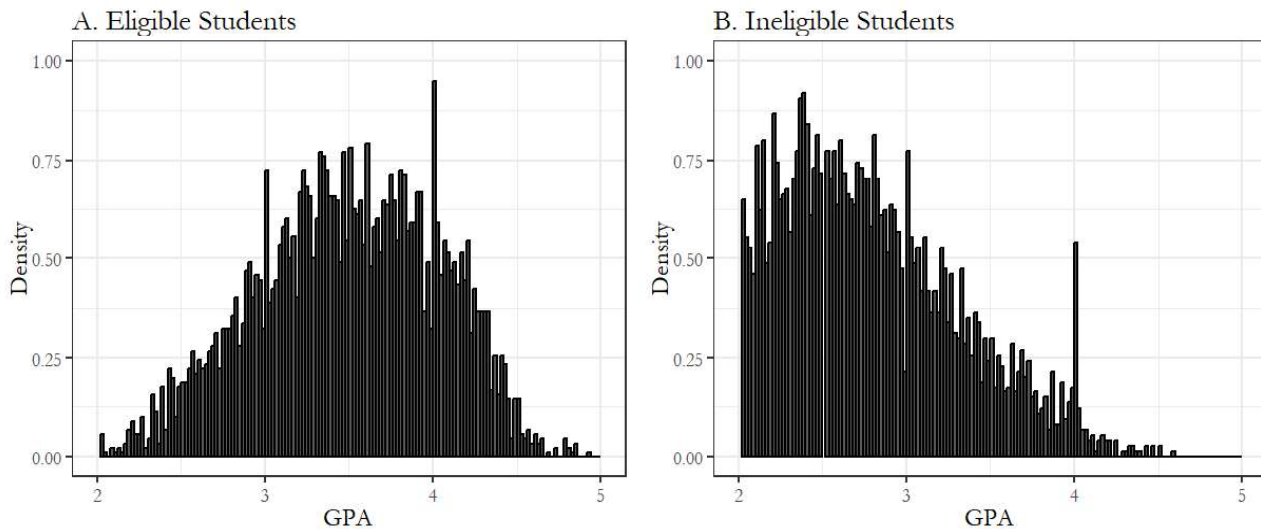


Figure 1. Distribution of Student GPAs. This figure presents histograms of the distribution of student GPAs using a bin size of 0.02. Eligible students is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only) and took the SAT I or ACT exams. Ineligible students is restricted to the remaining graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have not taken the SAT I or ACT exams or have not satisfied the A-G course requirement (VUSD only), thereby making them ineligible for the CSUSM MOUs. Students with a GPA less than 2.0 are not depicted in the graph.

However, the histogram distribution in **Figure 1** visually suggests a slight bunching just above the 3.0 GPA cutoff for both eligible (insignificant) and ineligible (significant at the one percent level) students. As Zimmerman (2014) argues,⁵⁰ bunching at the 3.0 GPA cutoff may be traced to other factors, making the standard test of distributional discontinuity unhelpful. For instance, a GPA of 3.0 corresponds to an average, unweighted grade of “B”, which is a benchmark grade level that teachers may be more likely to assign or students may be more likely to strive for. The jumps in the density at other non-cutoff GPA points (e.g., at a GPA of 4.0), especially for students who are ineligible for the CSUSM MOUs, suggest that these other factors may be important. Therefore, the author conducts an additional visual test following Zimmerman (2014),⁵⁰ which examines the continuity in the density ratios of the conditional densities to the unconditional density. The density ratios are presented in **Figure 2** for four different conditioning groups: first-generation, white, female, and low-income students. Consistent with a valid RD, there are no discontinuous jumps in density ratios around the 3.0 GPA cutoff.

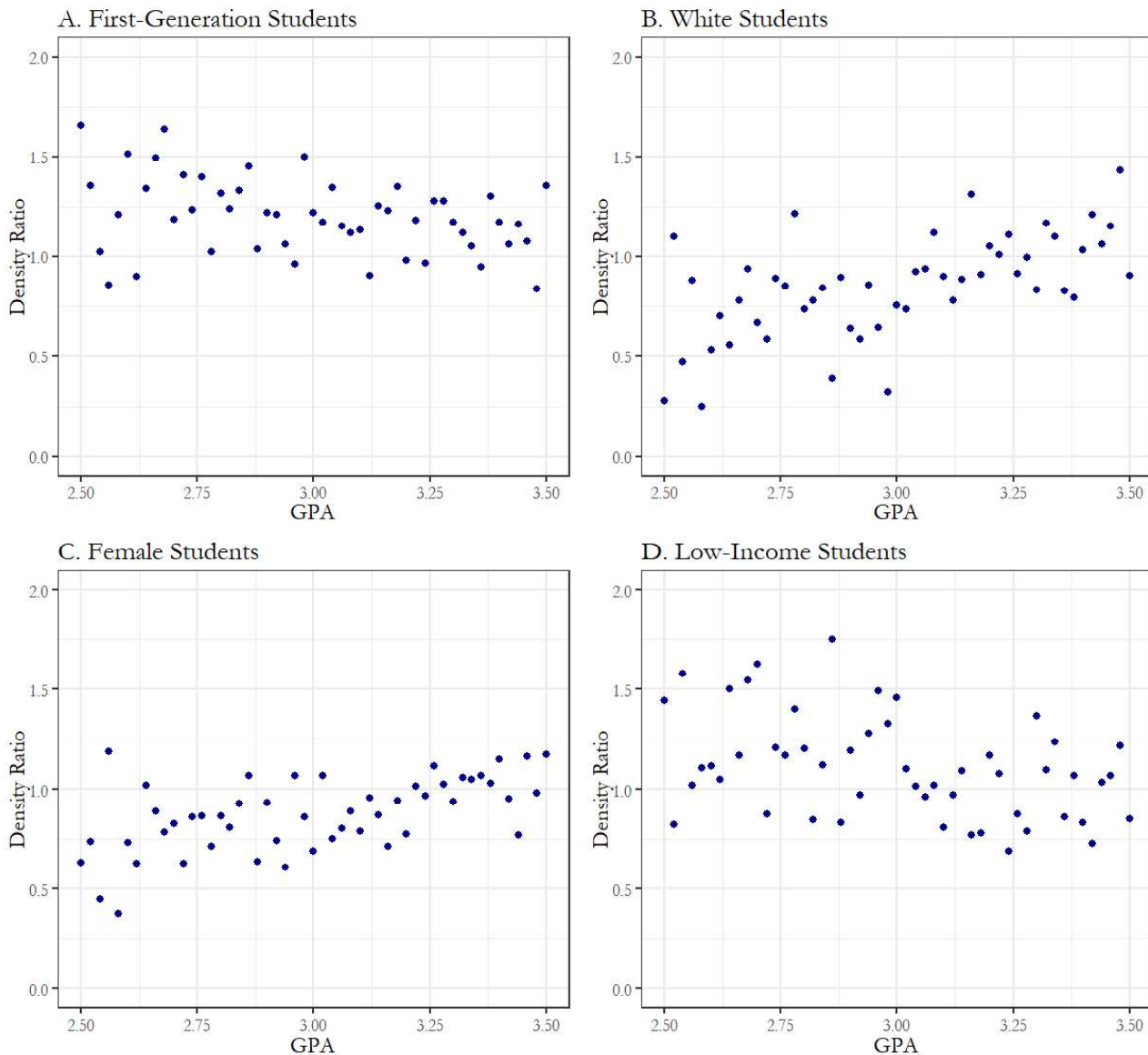


Figure 2. Density Ratios of Student GPAs. This figure presents ratios of the conditional to unconditional densities of student GPAs using a bin size of 0.02. Four different conditioning variables are used: first-generation, white, female, and low-income students. The sample is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only) and took the SAT I or ACT exams.

In addition, the author checks if observable covariates are balanced across the 3.0 GPA cutoff. **Figure 3** plots the proportion of students under each covariate across the 3.0 GPA cutoff and **Table 3** presents the corresponding linear regression estimates. There are no significant differences in first-generation and low-income status across the GPA cutoff, but there are higher proportions of white students and lower proportions of female students just above the GPA cutoff, suggesting that white and male students may be more likely to bargain for a 3.0 GPA in lieu of a slightly lower GPA.^Q

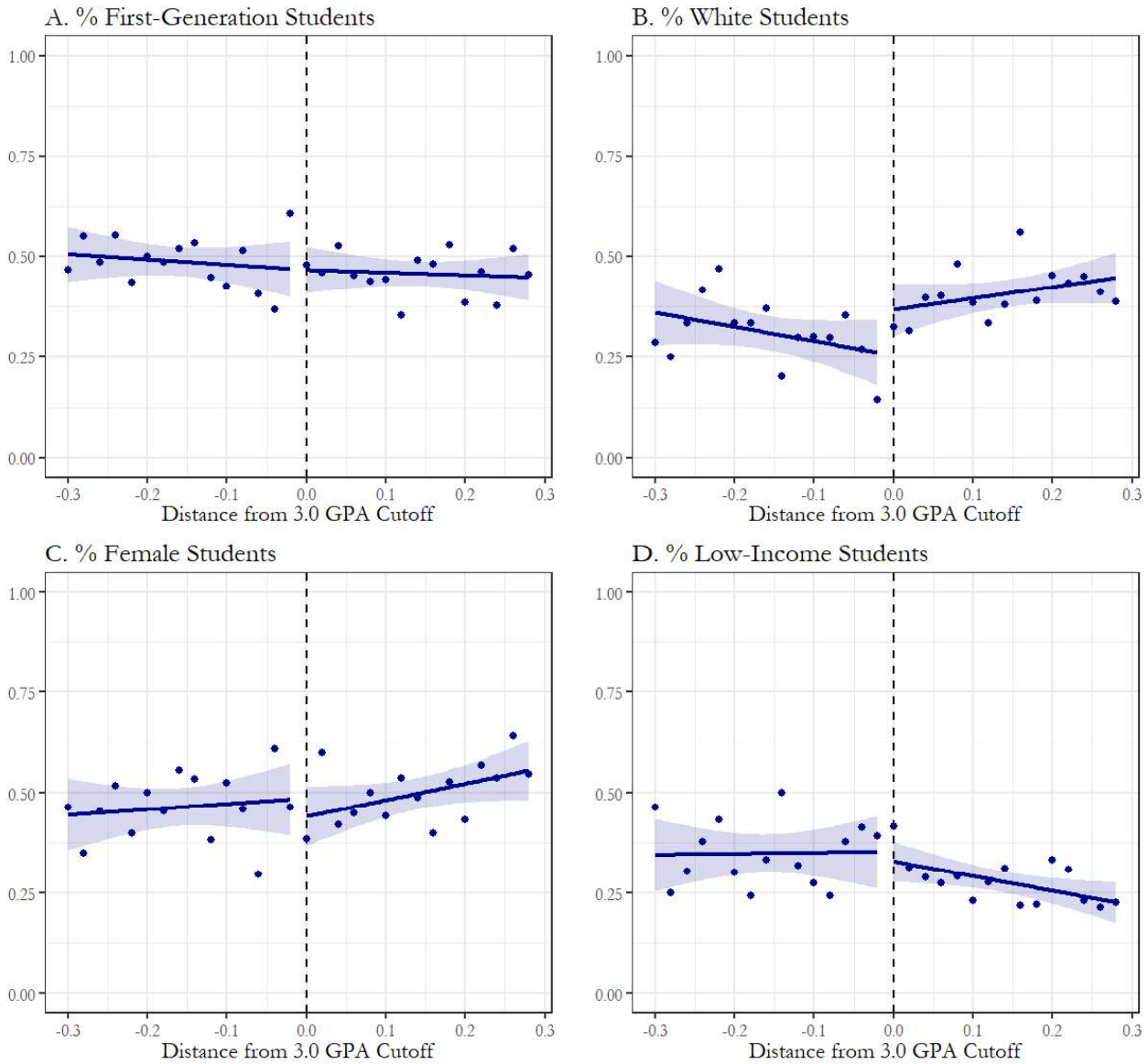


Figure 3. Covariate Balance. This figure presents binned scatterplots and linear best-fit lines of the proportions of demographic variables of MVUSD and VUSD graduates by their GPA distance to the 3.0 GPA cutoff using a bin size of 0.02. The sample is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only), took the SAT I or ACT exams, and have a GPA within 0.3 of the GPA cutoff.

	% First-Generation	% White	% Female	% Low-Income
	(1)	(2)	(3)	(4)
Above 3.0 GPA Cutoff	0.004107 (0.047872)	0.081321* (0.046659)	-0.08074* (0.047838)	-0.02792 (0.044102)
Control Mean	0.464209	0.31947176	0.51305048	0.3346032

Table 3. Covariate Balance (BW: [2.73, 3.43]). This table presents the estimated difference in the proportion of a demographic variable just above the 3.0 GPA cutoff using a linear regression. The sample is restricted to the RD sample (graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only), took the SAT I or ACT exam, and have a GPA between 2.73 to 3.43. *: p-value < .1; **: p-value < .05; ***: p-value < .01.

Lastly, the author conducts a placebo exercise using data from the graduating class of 2016 for MVUSD. Since the first MVUSD graduating class eligible for the CSUSM MOU was the graduating class of 2017, the class of 2016 should have experienced no discontinuous jumps in college enrollment around the 3.0 GPA cutoff. As expected, there are no significant increases in enrollment at any postsecondary, any four-year institution, CSUSM, or any CSU nor any significant decreases in enrollment at any two-year institutions for the placebo sample. Instead, prior to the implementation of the CSUSM MOUs, there is a slight (but insignificant) decrease in four-year and CSUSM enrollment and an increase in two-year enrollment across the 3.0 GPA cutoff (See **Appendix Table A1** and **Appendix Figure A2**).^R In contrast, the MVUSD graduating classes of 2017 to 2019 have similar enrollment trends compared to the full sample of MVUSD and VUSD graduates that will be outlined in **Table 4** (See **Appendix Table A2**).^S

RESULTS

Figure 4 plots each outcome variable with the distance from the 3.0 GPA cutoff as the running variable for the full RD sample. As expected, there does not appear to be a noticeable change in postsecondary enrollment rates around the 3.0 GPA cutoff in Panel A, due to the already high baseline college enrollment rate of 87 percent. Despite the little change in college enrollment, Panels B and C reveal a slight change in the type of institution students enrolled at. Above the 3.0 GPA cutoff, there is a 6.8 percentage point increase in four-year college enrollment, which is a 15 percent gain from the control mean, and a 5.9 percentage point decrease (14 percent from the control mean) in two-year enrollment, but both estimates are not statistically significant at conventional levels (See **Table 4**). These trends suggest that the MOUs encouraged students who were already enrolling in postsecondary institutions to shift away from two-year institutions towards four-year institutions. Panels D and E show that this increase in four-year enrollment is driven by increased enrollment at CSUSM and at the CSU system in general, where attendance increased by 6.1 percentage points at CSUSM (from a base of 15 percent) and by 7.7 percentage points at any CSU (from a base of 23 percent).^T

Heterogeneity by Baseline Sociodemographic Characteristics

Next, the paper analyzes if the baseline results differed by sociodemographic characteristics as guaranteed admission programs traditionally benefited underrepresented students the most after the banning of affirmative action in college admissions. Furthermore, nonwhite (in particular, Black and Hispanic) students, low-income students, and first-generation college students are more likely to enroll at two-year institutions compared to their counterparts, so they may be more likely to shift from two-year to four-year institutions under the CSUSM MOUs.^{14, 52} Moreover, the CSU system serves a more diverse community of underrepresented and low-income students compared to other four-year colleges, potentially making guaranteed admission to CSUSM more appealing to these groups of students.¹⁴

Breaking down the effects by baseline sociodemographic characteristics reveals that the significant increases in enrollment at CSUSM and at any CSU are driven mostly by first-generation, nonwhite, and low-income students. The results, summarized in columns two through nine of **Table 4**, are also presented in **Figures 5 – 8**. Similar to the main baseline results, first-generation, nonwhite, and low-income students see higher, but insignificant, four-year enrollment rates and lower, but insignificant (with the exception of first-generation students), two-year enrollment rates driven mostly by significant increases in CSUSM and CSU enrollment rates.

Figure 5 plots each enrollment outcome by first-generation status and columns two and three of **Table 4** present the corresponding regression estimates. While non-first-generation students see little impact on enrollment across all categories, first-generation students drive the results from the full sample. Below the 3.0 GPA cutoff, first-generation students are less likely to enroll at a four-year institution or at any CSU and more likely to enroll at a two-year compared to their non-first-generation peers. However, first-generation students just above the cutoff are 10.3 percentage points (21 percent gain from the control mean) more likely to enroll at a four-year institution (not statistically significant), 12.1 percentage points (56 percent) more likely to enroll at any CSU, and 12.2 percentage points (27 percent) less likely to enroll at any two-year institution. Thus, above the 3.0 GPA cutoff, the CSUSM MOUs closed the gap between first-generation and non-first-generation four-year college enrollment rates.

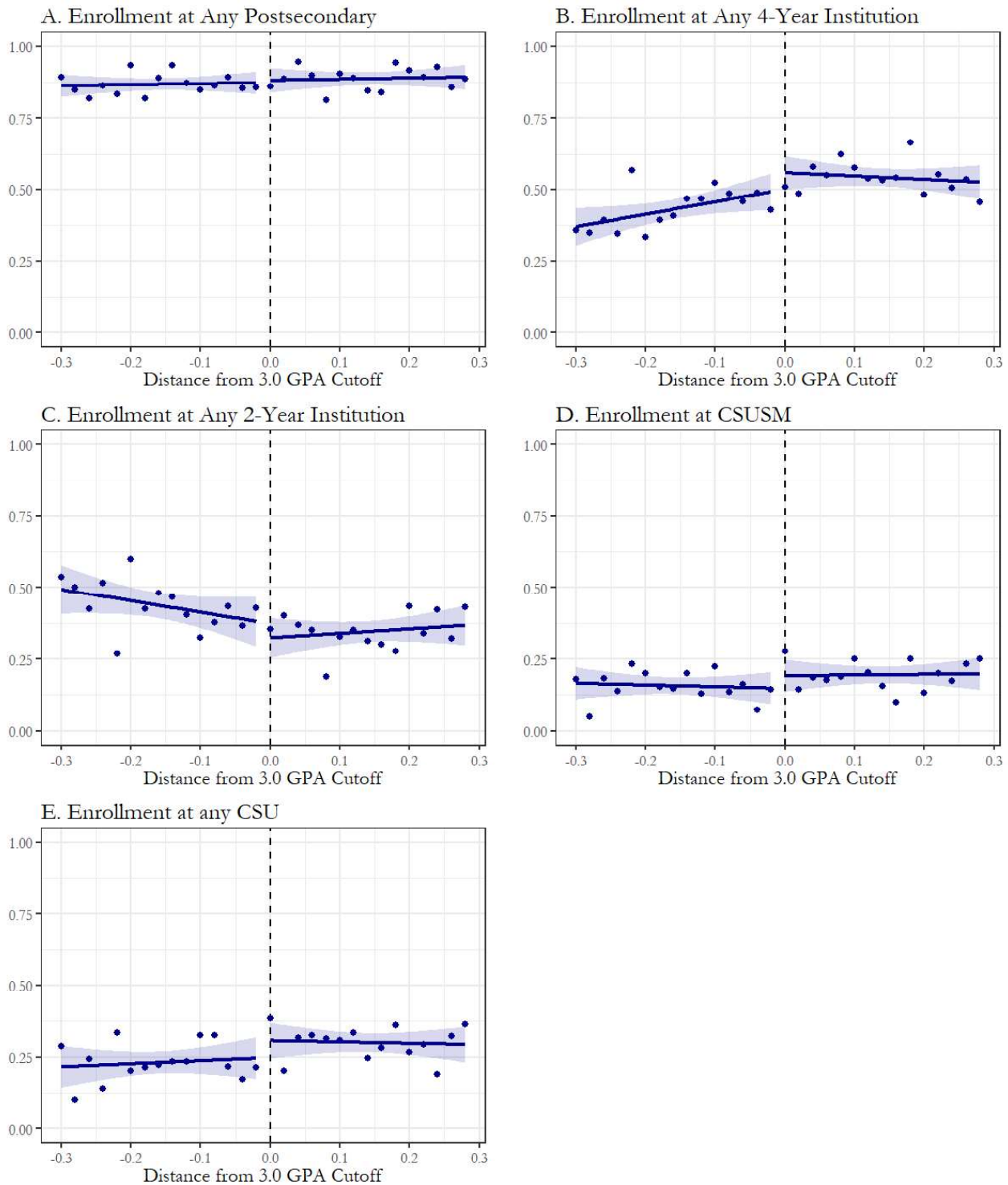


Figure 4. Enrollment Rates of the Full Sample. This figure presents binned scatterplots and linear best-fit lines of the enrollment outcomes of MVUSD and VUSD graduates by their GPA distance to the 3.0 GPA cutoff using a bin size of 0.02. The sample is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only), took the SAT I or ACT exams, and have a GPA within 0.3 of the GPA cutoff.

	By First-Gen Status			By Race		By Gender		By Low-Income Status	
	Full Sample (N = 1616)	First-Generation (N = 732)	Not First-Generation (N = 884)	Nonwhite (N = 991)	White (N = 625)	Female (N = 831)	Male (N = 785)	Low-Income (N = 488)	Not Low-Income (N = 1128)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Any Postsecondary									
RD Estimate	0.009261 (0.031122)	-0.01873 (0.047284)	0.032598 (0.041285)	0.011724 (0.040115)	0.001572 (0.049594)	-0.01418 (0.043893)	0.030348 (0.04437)	0.034573 (0.065405)	-0.00649 (0.033735)
Control Mean	0.86993956	0.8695248	0.8709822	0.86220374	0.88522314	0.89329966	0.84898232	0.7988356	0.90561328
P Identical	0.411437			0.876876		0.475795		0.534084	
Any Four-year									
RD Estimate	0.068212 (0.047858)	0.103219 (0.070697)	0.038916 (0.064942)	0.034722 (0.05942)	0.115682 (0.08143)	0.074746 (0.068827)	0.074994 (0.066751)	0.059694 (0.084008)	0.068264 (0.058143)
Control Mean	0.45698954	0.41204546	0.49576752	0.45602672	0.46385	0.47305728	0.43617686	0.4036199	0.48442624
P Identical	0.502932			0.421745		0.997935		0.933141	
Any Two-year									
RD Estimate	-0.05895 (0.046051)	-0.12195* (0.06888)	-0.00632 (0.061853)	-0.023 (0.057107)	-0.11411 (0.078447)	-0.08893 (0.06536)	-0.04465 (0.065086)	-0.02512 (0.080558)	-0.07476 (0.056201)
Control Mean	0.412953	0.45747834	0.3751978	0.406176	0.4213762	0.4202404	0.41280646	0.3952316	0.421193
P Identical	0.21079			0.347292		0.631321		0.614372	
CSUSM Only									
RD Estimate	0.061472* (0.03699)	0.06899 (0.057051)	0.052972 (0.048159)	0.108597** (0.048458)	-0.01843 (0.056089)	0.074944 (0.055355)	0.05312 (0.049402)	0.123677* (0.067169)	0.030105 (0.044283)
Control Mean	0.1503784	0.15321706	0.149754	0.1496738	0.155076	0.1662956	0.1343166	0.1238444	0.1653666
P Identical	0.828854			0.101244		0.768596		0.236286	
Any CSU									
RD Estimate	0.077419* (0.043302)	0.120664* (0.064431)	0.039717 (0.058508)	0.133593** (0.055366)	-0.01911 (0.068772)	0.090178 (0.064733)	0.078822 (0.057648)	0.147789** (0.074913)	0.04079 (0.053031)
Control Mean	0.2332672	0.2169588	0.2480834	0.2324222	0.23992902	0.2586636	0.2040666	0.1793032	0.2618932
P Identical	0.351837			0.091993		0.89574		0.247736	

Table 4. RD Estimates by Type of College and Demographic (BW: [2.73, 3.43]). This table presents RD estimates of the effect of eligibility for the CSUSM MOUs on college enrollment rates by postsecondary institution type and demographic following Equation 1. The sample is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only) and the SAT/ACT completion requirement. Control mean reports estimated baseline enrollment rates for the bin (size = 0.02) below the 3.0 GPA cutoff, calculated as $\beta_0 + \beta_1 \cdot 2.98$ derived from Equation 1. P Identical represents p-values for the hypothesis that the coefficients between each pair of demographic groups are identical. For instance, the P Identical value in column two represents the hypothesis that first-generation and non-first-generation students have the same RD estimate. The bandwidth is calculated following Calonico et al (2014) and fixed from running the regression for the full sample of students and the outcome of enrollment at any postsecondary. *; **, ***; p-value < .05; **, ***; p-value < .01.

Plotting the enrollment outcomes by race reveals there is little difference in the enrollment rates at CSUSM and any CSU of nonwhite and white students before the 3.0 GPA cutoff (See **Figure 6**). However, just above the cutoff, nonwhite students are 10.9 percentage points (73 percent) more likely to enroll at CSUSM and 13.4 percentage points (57 percent) more likely to enroll at any CSU (both of which are significant at the five percent level), whereas white students are not significantly affected by the CSUSM MOUs across all enrollment outcomes (See columns four and five of **Table 4**). This result suggests that although the CSUSM MOUs are not increasing the overall postsecondary or four-year enrollment rates of nonwhite students, the program is changing the type of four-year institutions they are enrolling at towards the CSUs, in particular, towards CSUSM.

Breaking down the effects by gender reveals that the CSUSM MOUs did not affect female and male students differently. **Figure 7**, which plots each enrollment outcome by gender, shows that female students are more likely to enroll at any four-year institution, CSUSM, and any CSU than male students and this trend continued above the 3.0 GPA cutoff. In addition, both female and male students see a general increase in enrollment at four-year institutions, CSUSM, and any CSU and a decrease in enrollment at two-year institutions, but neither group experience significant effects (See columns 6 and 7 of **Table 4**).

Lastly, when breaking down the effects by income, low-income students experience greater enrollment impacts compared to students from higher-income backgrounds. **Figure 8**, which plots each enrollment outcome by income, reveals that low-income students are less likely to enroll at any postsecondary, any four-year institution, at CSUSM, and any CSU compared to non-low-income students below the 3.0 GPA cutoff. Just above the cutoff, low-income students are 12.4 percentage points (100 percent) more likely to enroll at CSUSM and 14.8 percentage points (82 percent) more likely to enroll at any CSU (See columns 8 and 9 of **Table 4**). Thus, the CSUSM MOUs doubled the likelihood of low-income students enrolling at CSUSM and nearly doubled the likelihood of low-income students enrolling at any CSU, closing the income enrollment gap at CSUSM and at CSUs. In addition, both low-income and higher-income students experienced a positive, but insignificant increase in four-year enrollment and a negative, but insignificant decrease in two-year enrollment.

Robustness

To verify the robustness of the results, the author runs the analysis with a larger sample of students by increasing the bandwidth size to 0.6 and 0.7 above and below the 3.0 GPA cutoff (roughly twice as large as the data-driven bandwidths) and including the graduating class of 2020. The author also plots each enrollment outcome using a fourth-order polynomial line of best fit (compared to the linear lines of best fits presented above).

When the bandwidth size is increased to 0.6, the estimates all move in the same direction as the main results, though some estimates are insignificant at the 10% level (See **Appendix Table A5**). When the bandwidth size is increased to 0.7, the estimates once again move in the same direction as the main results and more estimates become significant at the 10% level compared to the data-driven bandwidth (See **Appendix Table A6**). Hence, the positive trends in four-year, CSUSM, and CSU enrollment rates and the negative trend in two-year enrollment just above the GPA cutoff are robust to bandwidth size, but the significance levels of the estimates may vary due to the sample size.

When including the graduating class of 2020 in the analysis, the data-driven bandwidth increases slightly to a GPA between 2.74 to 3.44 (a 0.01 increase above and below the 3.0 cutoff). In this new bandwidth, the estimates all move in the same direction as the main results, though estimates for the full sample of students and low-income students become not significant (See **Appendix Table A7**). If the bandwidth is adjusted to 0.6, there are no changes in the direction of the results, but if the bandwidth is increased to 0.7, more estimates become significant compared to the data-driven bandwidths, similar to the effects of changing bandwidths in the original RD sample (See **Appendix Tables A8 – A9**). This suggests that the graduating class of 2020 experienced similar impacts from the CSUSM MOUs as previous

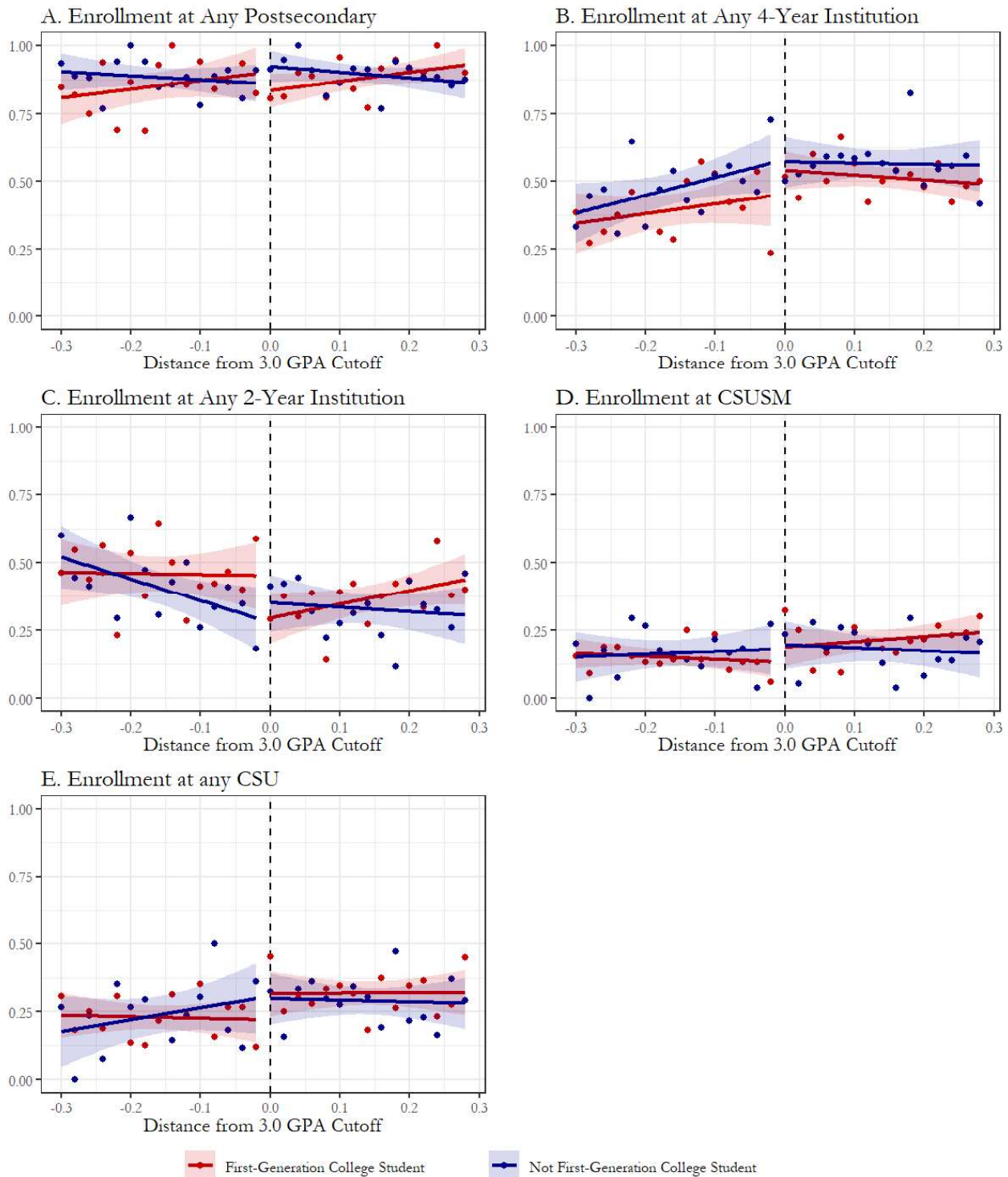


Figure 5. Enrollment Rates by Parents' Educational Attainment. This figure compares binned scatterplots and linear best-fit lines of the enrollment outcomes of first-generation MVUSD and VUSD graduates with non-first-generation graduates by their GPA distance to the 3.0 GPA cutoff using a bin size of 0.02. The sample is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only), took the SAT I or ACT exams, and have a GPA within 0.3 of the GPA cutoff.

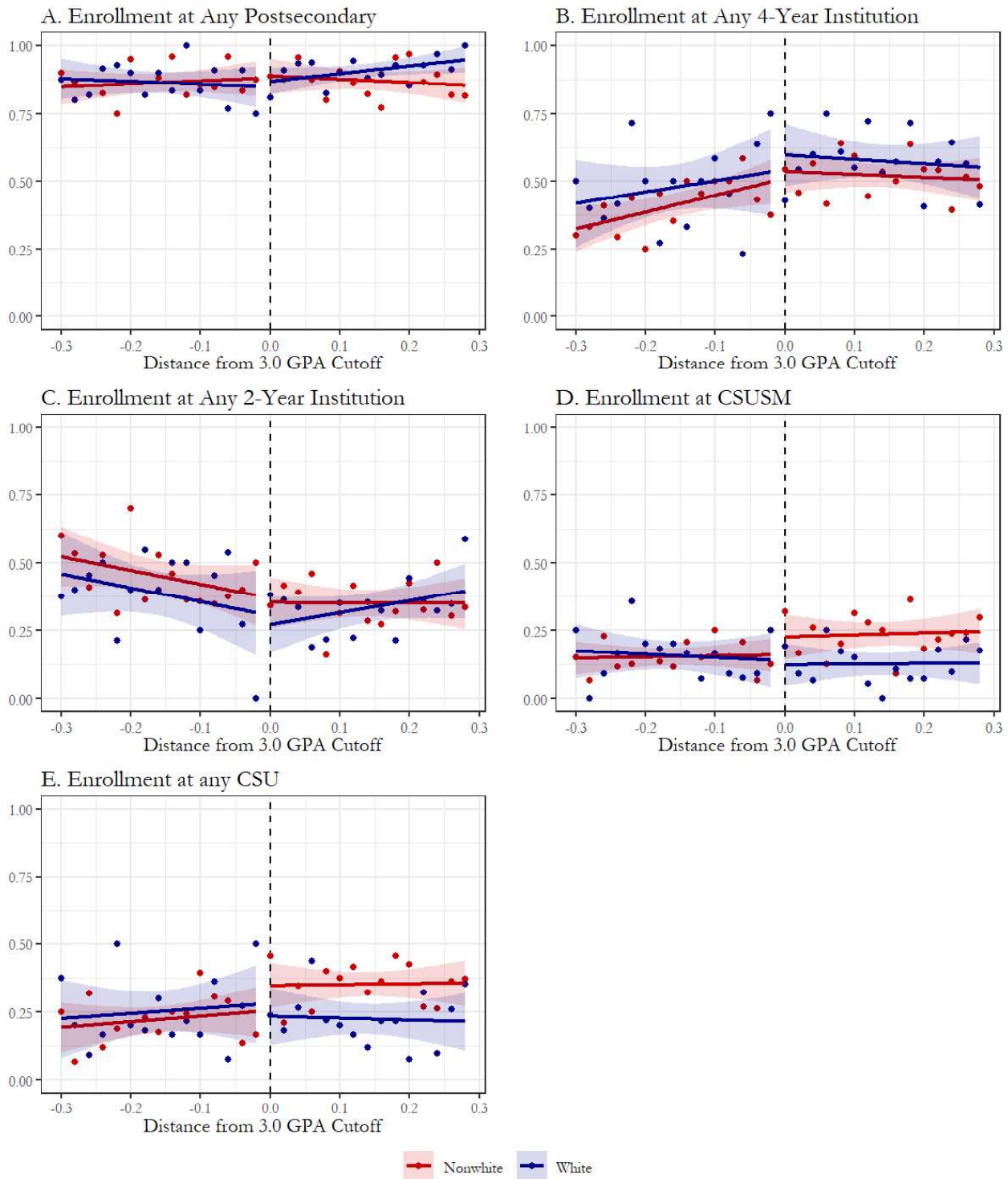


Figure 6. Enrollment Rates by Race. This figure compares binned scatterplots and linear best-fit lines of the enrollment outcomes of nonwhite MVUSD and VUSD graduates with white graduates by their GPA distance to the 3.0 GPA cutoff using a bin size of 0.02. The sample is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only), took the SAT I or ACT exams, and have a GPA within 0.3 of the GPA cutoff.

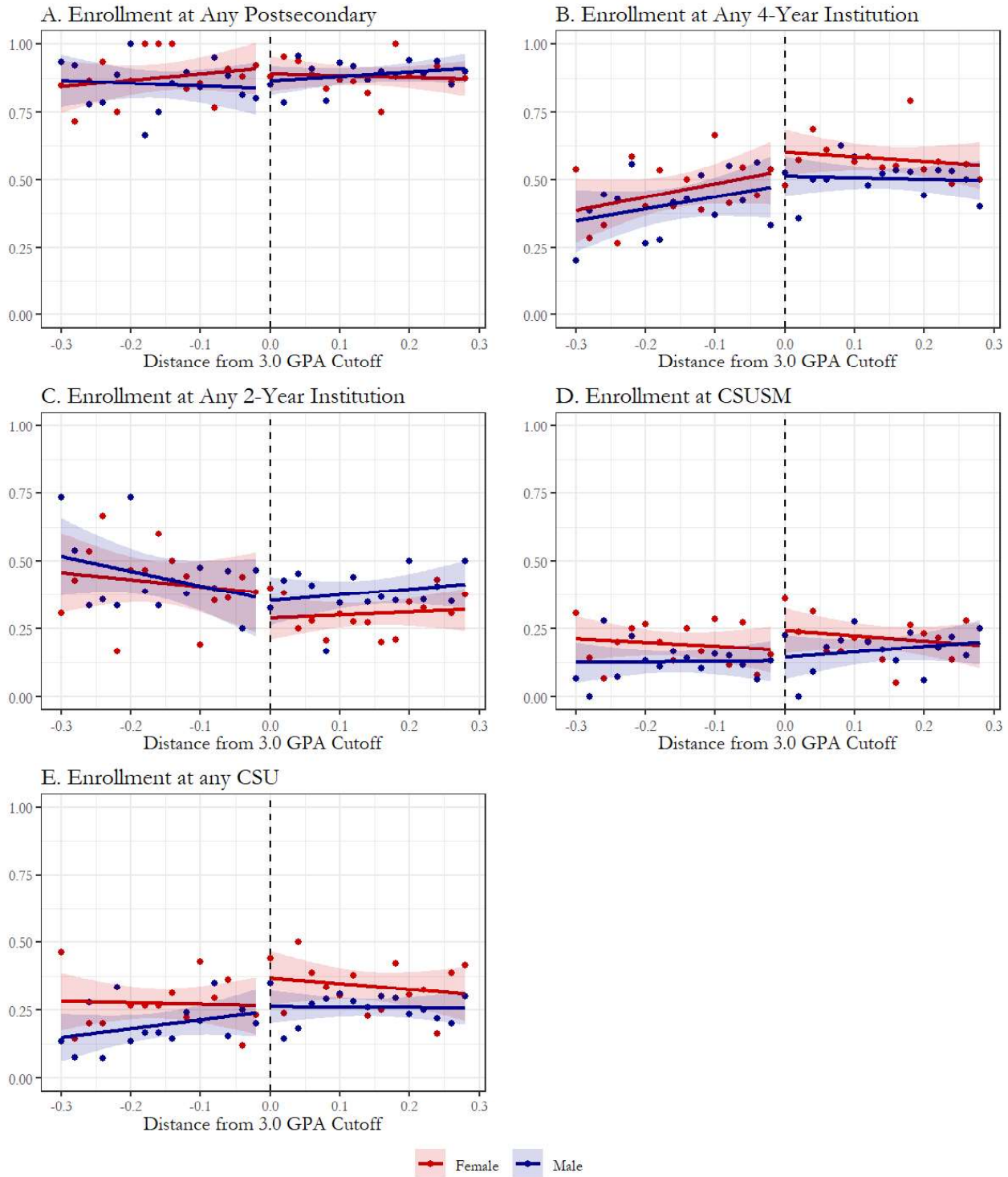


Figure 7. Enrollment Rates by Gender. This figure compares binned scatterplots and linear best-fit lines of the enrollment outcomes of female MVUSD and VUSD graduates with male graduates by their GPA distance to the 3.0 GPA cutoff using a bin size of 0.02. The sample is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only), took the SAT I or ACT exams, and have a GPA within 0.3 of the GPA cutoff.

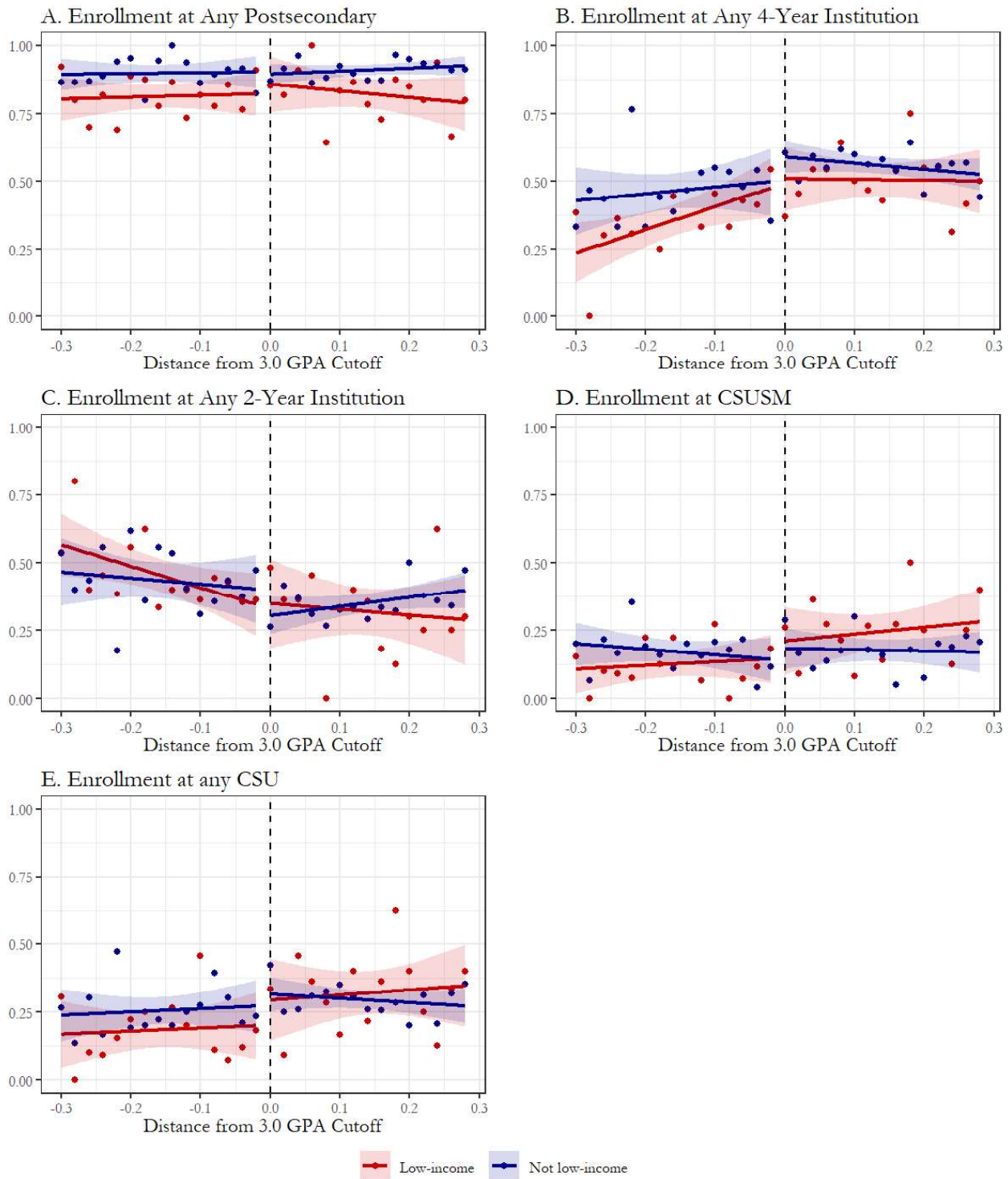


Figure 8. Enrollment Rates by Income. This figure compares binned scatterplots and linear best-fit lines of the enrollment outcomes of low-income MVUSD and VUSD graduates with non-low-income graduates by their GPA distance to the 3.0 GPA cutoff using a bin size of 0.02. The sample is restricted to graduates from 2017 to 2019 for MVUSD and 2018 to 2019 for VUSD who have satisfied the A-G course requirement (VUSD only), took the SAT I or ACT exams, and have a GPA within 0.3 of the GPA cutoff.

graduating classes but potentially at a smaller magnitude due to uncertainty surrounding COVID-19 and online learning.

Lastly, the author provides plots of enrollment outcomes using a fourth-order polynomial line of best fit for the full sample and breakdowns by heterogeneity (See **Appendix Figures A3 – A7**). Similar to the main results, the plots depict significant increases in CSUSM and CSU enrollment rates for the full sample, driven mostly by first-generation, non-white, and low-income students. In addition, the figures demonstrate that the CSUSM MOUs close the college enrollment gaps between first-generation and non-first-generation students along with the income enrollment gaps at CSUSM and CSUs.

DISCUSSION AND CONCLUSION

The Memorandums of Understanding between California State University San Marcos and its local school districts differ from previously studied state-wide guaranteed admission programs as they mainly targeted students who were considering two-year options to enroll at a guaranteed, local four-year option instead. This paper finds that the CSUSM MOUs primarily affected underrepresented students, with first-generation, nonwhite, and low-income students experiencing the greatest increase in CSUSM and CSU enrollment along with a shift away from two-year enrollment to four-year enrollment (not statistically significant) under the MOUs. This suggests that local guaranteed admission programs like the CSUSM MOUs can also serve as a force for increasing the enrollment rates of underrepresented students at higher-quality, public four-year institutions like previously studied state-wide guaranteed admission programs.

However, it is important to consider the limitations of the research. This paper focused on the two districts from which the author could obtain data, but CSUSM has signed MOUs with eight other districts. With richer data, it may be possible to gain insight into how proximity to the college could also impact the effectiveness of a local guaranteed admission program and obtain more precise estimates thanks to the greater sample size. In addition, since the CSUSM MOUs are a relatively new guaranteed admission program, the author is unable to observe long-run outcomes such as college graduation rates and earnings after high school for most graduation cohorts. Future research can follow students and observe their long-run outcomes to better understand if the increased enrollment at CSUSM and any CSU resulted in any “mismatch” of underrepresented students.

On the flip side of the mismatch theory, an additional potential concern from these findings is whether students would have continued enrolling at 2-year institutions and transferred to higher-quality 4-year institutions if not for the MOUs. Although the author does not have data on institutions students transferred to (if any) and long-run outcome data, transfer rates remain low on average with only 2-3 percent of California community college students transferring to a 4-year institution within 2 years and 22-24 percent transferring within 4 years.⁵³ Average U.S. 2-year college graduation rates within 4 years (34 percent) are also lower than average U.S. 4-year college graduation rates within 4 years (64 percent).⁵⁴ Thus, the author argues that the CSUSM MOUs as a whole likely help local students gain access to 4-year colleges rather than steer them away from higher-quality options.

Despite the limitations of this research, the analysis provides insight into the effectiveness of guaranteed admission programs and how the requirements of each program may shape its effectiveness. Notably, this paper did not observe any changes to enrollment at any postsecondary as a result of the CSUSM MOUs, likely due to already high baseline college enrollment rates when conditioning on the sample of students who have completed the SAT/ACT exam and the A-G requirements. This indicates that in order for college access programs to be effective in expanding overall college enrollment, not just enrollment at public four-year institutions or higher-quality institutions, they should include students who have not already exhibited an interest in higher education. This implication may be particularly interesting in the post-COVID-19 world, as many colleges are moving away from requiring standardized exam scores like the SAT/ACT, allowing students to more easily apply to colleges and utilize college access programs like guaranteed admission programs. However, one common requirement (which is also affecting future graduating classes under the

CSUSM MOUs) that can continue to restrict students' participation in college access programs and consequently also restrict the programs' gains in college enrollment is for students to complete the Free Application for Federal Student Aid (FAFSA). Thus, future policymakers and researchers may be interested in understanding how to increase college access in conjunction with raising awareness of financial aid opportunities.

ACKNOWLEDGEMENTS

The author thanks the administrative teams at Murrieta Valley Unified School District and Vista Unified School District for the time and effort they put into preparing student data for this research.

NOTES

- A. Although the MOUs were signed between 2009 to 2015, the first graduating classes to be affected by this program in each district ranged from 2009 to 2019.
- B. 9,032 students graduated from the two school districts between 2017 to 2019, and roughly half satisfied the A-G course requirement and SAT I/ACT requirement.
- C. The TTP was adjusted in 2009 to place a cap on the share of first-year in-state students who are guaranteed admission to UT Austin to 75 percent of the incoming class.¹⁹ The findings on TTP mentioned in this paper all refer to impacts before 2009.
- D. The ELC was expanded to students graduating in the top nine percent of their California high school class in 2012. As a result, ELC-eligible students no longer experienced admission advantages at selective ("Absorbing") UCs.¹¹ The impacts described in this paper occurred before 2012.
- E. Enrollment numbers are calculated from the Integrated Postsecondary Education Data System (IPEDS), which is provided by the U.S. Department of Education, National Center for Education Statistics.
- F. Underrepresented minority is defined in this paper as a student who identifies as Black or African American, Hispanic, Native American or Alaskan Native, Pacific Islander, or as two or more races.
- G. CSUs no longer require students to provide SAT or ACT scores as of March 2022, but SAT or ACT scores were still required during the years studied in this paper.²²
- H. Beginning with the 2021-2022 admissions cycle, the CSU application asks some questions about extracurricular involvement. Additionally, California Polytechnic State University, San Luis Obispo asked supplemental questions on students' extracurricular involvement prior to this change.^{16, 17}
- I. A map of the local area CSUSM serves is provided in the appendix, along with the school districts that signed an MOU with CSUSM and the districts specifically studied in this paper (See **Appendix Figure A1**).
- J. CSUSM now requires students to complete the Free Application for Federal Student Aid (FAFSA) to receive guaranteed admission, but the eligibility conditions are based on the most recent edition of each school district's signed MOU with CSUSM at the time of data collection and analysis in 2022.²⁵
- K. Beginning with the graduating class of 2022, the EPT and ELM testing were eliminated and completely replaced by other placement evaluations like SAT, ACT, and Advanced Placement (AP) exam scores.²⁶
- L. Students who graduated from San Marcos Unified School District did not need to satisfy the GPA condition to receive guaranteed admission to CSUSM.²⁸

- M. Although the first VUSD graduating class affected by the CSUSM MOUs was the graduating class of 2017, the first year of data observed starts with the class of 2018.
- N. Although the graduating class of 2020 was not impacted by the COVID-19 pandemic until the second half of the school year, they still may have been affected by the uncertainty surrounding online learning. This paper experiments with including the class of 2020 in the analysis in the later “Robustness” section.
- O. When running the analysis without conditioning on VUSD students satisfying the A-G requirements, the estimates become larger because the average college enrollment rates below the 3.0 GPA cutoff become lower. However, this is not a perfect indication that not conditioning on MVUSD students satisfying the A-G requirements is overestimating the effect since MVUSD students have higher average GPAs and are more likely to complete the A-G requirements compared to VUSD students. Thus, the distribution of MVUSD students who did not complete the A-G requirements may be more equal on both sides of the GPA cutoff. The results without conditioning on VUSD students satisfying the A-G requirement are available upon request.
- P. First-generation college student is defined in this paper as a student whose parents’ highest level of educational attainment is some college or less (did not acquire a four-year degree).
- Q. This tendency for male students to be more likely than female students to bargain for better grades has been documented at the university level.⁵¹
- R. One exception is that white students see a significant decrease in four-year enrollment and a significant increase in two-year enrollment above the 3.0 cutoff. However, these results move in the opposite direction of the main results presented in the next section, suggesting that white students who are more likely to “manipulate” their GPA to be above 3.0 are also more likely to attend two-year colleges and not four-year colleges.
- S. The author experiments with matching bandwidths with the full RD sample (BW: 2.73 to 3.43) along with using data-driven bandwidths outlined by Calonico *et al.* (2014) for the enrollment outcome of any postsecondary for both the placebo and the MVUSD graduating classes of 2017 to 2019 samples.⁴⁸ Both approaches yield very similar bandwidths and hence, little difference in estimates (See **Appendix Tables A3 and A4**).
- T. The author also considers if the CSUSM MOUs affect UC enrollment rates, potentially by encouraging students to choose a local option rather than a UC. There are no observed impacts on UC enrollment due to an overall low UC enrollment rate around the 3.0 GPA cutoff. The median GPA of students who enrolled at any UC in the full sample was 4.086, and only 23 students in the RD sample enrolled at a UC.

REFERENCES

1. Torpey, E., Education level and projected openings, 2019-29, <https://www.bls.gov/careeroutlook/2020/article/education-level-and-openings.htm#:~:text=A%20bachelor's%20degree%20is%20typically,average%\protect\T1\textbraceright2C%20from%202019%20to%202029> (accessed May 2023)
2. U.S. Bureau of Labor Statistics, Education pays, 2021, <https://www.bls.gov/careeroutlook/2022/data-on-display/education-pays.htm> (accessed May 2023)
3. U.S. Bureau of Labor Statistics, High school graduates with no college had unemployment rate of 4.5 percent in February 2022, <https://www.bls.gov/opub/ted/2022/high-school-graduates-with-no-college-had-unemployment-rate-of-4-5-percent-in-february-2022.htm> (accessed May 2023)
4. U.S. Department of Education, Office of Planning, Evaluation and Policy Development and Office of the Under Secretary, Advancing Diversity and Inclusion in Higher Education: Key Data Highlights Focusing on Race and Ethnicity and Promising Practices, <https://www2.ed.gov/rschstat/research/pubs/advancing-diversity-inclusion.pdf>

(accessed May 2023)

5. Bailey, M. J., and Dynarski, S. M. (2011) Inequality in Postsecondary Education, in *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances* (Duncan G. J., and Murnane, R. J.), 117–132, Russell Sage Foundation, New York.
6. Hoxby, C., and Avery, C. (2013) The Missing “One-Offs”: The Hidden Supply of High-Achieving, Low-Income Students, *Brookings Pap Econ Act* 2013 (1), 1–65. <http://doi.org/10.1353/eca.2013.0000>
7. Perna, L. W., Rowan-Kenyon, H., Bell, A., Thomas, S. L., and Li, C. (2008) A Typology of Federal and State Programs Designed to Promote College Enrollment, *J Higher Educ* 79(3), 243–267. <https://doi.org/10.1080/00221546.2008.11772098>
8. Black, S. E., Denning, J. T., and Rothstein, J. (2023) Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes, *Am Econ J: Appl Econ* 15(1), 26–67. <https://doi.org/10.1257/app.20200137>
9. Bleemer, Z. (2019) Diversity in University Admissions: Affirmative Action, Percent Plans, and Holistic Review, CSHE.6.19, <https://escholarship.org/uc/item/1kb1b4cq> (accessed May 2023)
10. Callahan, J., States with Automatic or Guaranteed College Admissions Policies, <https://www.cga.ct.gov/2021/rpt/pdf/2021-R-0077.pdf> (accessed December 2022)
11. Bleemer, Z. (2021) Top Percent Policies and the Return to Postsecondary Selectivity, CSHE.1.21, <https://escholarship.org/uc/item/6qf1888s> (accessed May 2023)
12. Odle, T. K., and Delaney, J. A. (2022) You are Admitted! Early Evidence on Enrollment from Idaho's Direct Admissions System, *Res High Educ* 63(6), 899–932. <https://doi.org/10.1007/s11162-022-09675-x>
13. The California State University, Access and Impaction at the CSU, <https://www.calstate.edu/attend/impaction-at-the-csu#:~:text=CSU%20Local%20Admission%20%26%20Service%20Areas,university%20are%20given%20priority%20admission> (accessed December 2022)
14. Rodriguez, O., Mejia, M. C., and Johnson, H. (2017) Increasing Equity and Diversity, https://www.ppic.org/wp-content/uploads/r_0917orr.pdf (accessed May 2023)
15. The California State University, Freshman: Admission Requirements, https://www.calstate.edu/apply/freshman/getting_into_the_csu/pages/admission-requirements.aspx (accessed March 2023)
16. Cal State Apply, Freshman Application Guide 2021-2022 <https://www.calstate.edu/apply/freshman/Documents/freshman-application-guide-21-22.pdf> (accessed May 2023)
17. Silbernagel, M. (2022) Applying to a Cal State University in 2022, <https://www.touchstoneadvising.com/post/applying-to-a-cal-state-university-in-2020> (accessed May 2023)
18. Cortes, K. E., and Lincove, J. A. (2018) Match or Mismatch? Automatic Admissions and College Preferences of Low- and High-Income Students, *Educ Eval Policy Anal* 41(1), 98–123. <https://doi.org/10.3102/0162373718813360>
19. Long, M. C., Saenz, V., and Tienda, M. (2010) Policy Transparency and College Enrollment: Did the Texas Top Ten Percent Law Broaden Access to the Public Flagships?, *Ann Am Acad Pol Soc Sci* 627(1), 82–105. <https://doi.org/10.1177/0002716209348741>
20. Fletcher, J. M., and Mayer, A. (2013) Tracing the Effects of Guaranteed Admission Through the College Process: Evidence from a Policy Discontinuity in the Texas 10% Plan, *Contemp Econ Policy* 32(1), 169–186. <https://doi.org/10.1111/coep.12011>
21. University of California, Subject Requirement (A-G), <https://admission.universityofcalifornia.edu/admission-requirements/freshman-requirements/subject-requirement-a-g.html> (accessed December 2022)
22. The California State University, High School Guidance - First-Time First-Year Students, <https://www.calstate.edu/apply/freshman/Pages/first-time-freshman-guidance.aspx> (accessed December 2022)
23. California State University San Marcos, Local Admission Area School Districts, <https://www.csusm.edu/admissions/how-to-apply/impaction/local-service-area.html> (accessed December 2022)
24. The California State University, Impacted Undergraduate Majors and Universities, 2023-24, <https://www.calstate.edu/attend/degrees-certificates-credentials/Pages/impacted-degrees.aspx> (accessed March 2023)

25. California State University San Marcos, K-12 Partner School Districts and Families, <https://www.csusm.edu/alliance/partners/index.html> (accessed December 2022)
26. California State University San Marcos, Readiness and Success Service: Early Start Exemptions, <https://www.csusm.edu/readiness/earlystart/exempt.html> (accessed March 2023)
27. Haynes, K. S., Boyle, S., Walters, J. and Hargrave, S. (2013) Memorandum of Understanding Between the Escondido Union High School District, Escondido Union School District, San Pasqual Union School District and California State University San Marcos, <https://www.csusm.edu/alliance/documents/a101625-eusd-eusd-spUSD-guaranteed-admission-open-ended.pdf> (accessed December 2022)
28. Haynes, K. S., and Holt, K. D. (2014) Memorandum of Understanding Between the San Marcos Unified School District and California State University San Marcos, <https://www.csusm.edu/alliance/documents/a101646-smusd-mou.pdf> (accessed December 2022)
29. Haynes, K. S., and Kelley, P. (2014) Memorandum of Understanding Between the Murrieta Unified School District and California State University San Marcos, <https://www.csusm.edu/alliance/documents/a101698-signed-mou-5.8.14.pdf> (accessed December 2022)
30. Haynes, K. S., and Kimberly, D. (2015) Memorandum of Understanding Between Lake Elsinore Unified School District and California State University San Marcos, <https://www.csusm.edu/alliance/documents/a102274-lake-elsinore-usd-guaranteed-admission-open-ended.pdf> (accessed December 2022)
31. Haynes, K. S., and Lovely, S. (2013) Memorandum of Understanding Between the Carlsbad Unified School District and California State University San Marcos, https://www.csusm.edu/alliance/documents/a101584-carlsbad-usd-guaranteed-admission-6_30_17.pdf (accessed December 2022)
32. Haynes, K. S., and Obermeyer, L. (2013) Memorandum of Understanding Between the Valley Center-Pauma Unified School District and California State University San Marcos, <https://www.csusm.edu/alliance/documents/a101594-valley-center-pauma-unified-school-district-guaranteed-admission-open-ended.pdf> (accessed December 2022)
33. Haynes, K. S., O'Neal, V., and Ritter, T. (2013) Memorandum of Understanding Between the Temecula Valley Unified School District and California State University San Marcos, https://www.csusm.edu/alliance/documents/a101604-tvUSD-guaranteed-admission-11_15_13.pdf (accessed December 2022)
34. Haynes, K. S., Pedroza, H. A., Singh, C., Koeppen, M. (2018) Memorandum of Understanding Between the Fallbrook Union High School District, Fallbrook Union Elementary School District, Vallecitos School District and California State University San Marcos, <https://www.csusm.edu/alliance/documents/2018/2018fubsdamendedmou.pdf> (accessed December 2022)
35. Haynes, K. S., and Perondi, L. (2014) Memorandum of Understanding Between Oceanside Unified School District and California State University San Marcos, https://www.csusm.edu/alliance/documents/2017/a101617-oceanside-unified-school-district-guaranteed-admission-6_30_17.pdf (accessed December 2022)
36. Haynes, K. S., and Vodicka, D. (2013) Memorandum of Understanding Between the Vista Unified School District and California State University San Marcos, <https://www.csusm.edu/alliance/documents/a101532-vUSD-signed-mou.pdf> (accessed December 2022)
37. Carlsbad High School, CHS Graduation Requirements, https://carlsbadhs.schoolloop.com/pf4/cms2/view_page?d=x&group_id=1587874782071&vdid=i13vb2hbodpuz677 (accessed May 2023)
38. El Camino High School, Graduation Requirements, <https://www.oside.us/domain/728> (accessed May 2023)
39. Elsinore High School, Graduation Requirements, https://ehs.leusd.k12.ca.us/apps/pages/index.jsp?uREC_ID=323677&type=d&pREC_ID=731744 (accessed May 2023)
40. Escondido Union High School District, EUHSD Graduation Requirements, https://www.eusd.org/apps/pages/index.jsp?uREC_ID=2757991&type=d&pREC_ID=2299361 (accessed May 2023)
41. Fallbrook Union High School, Graduation Requirements, <https://www.fallbrookhs.org/fallbrook/graduation-requirements> (accessed May 2023)
42. Murrieta Valley Unified School District, Graduation Requirements, <https://www.murrieta.k12.ca.us/Page/25321> (accessed May 2023)

43. San Marcos Unified School District, 2020-2021 Course Catalog, <https://sanmarcoshigh.smusd.org/common/pages/UserFile.aspx?fileId=30967620> (accessed May 2023)
44. Temecula Valley Unified School District, High School Graduation Requirements & College Entrance Requirements, <https://www.tvusd.k12.ca.us/Page/22991> (accessed May 2023)
45. Valley Center-Pauma Unified School District, Board Policy, https://core-docs.s3.amazonaws.com/documents/asset/uploaded_file/2756/VCPUSD/2218493/BP_6146.1_High_School_Graduation_Requirements.docx.pdf (accessed May 2023)
46. Vista Unified School District, High School Graduation Requirements, <https://www.vistausd.org/common/pages/DisplayFile.aspx?itemId=21140797> (accessed May 2023)
47. Riverside County Office of Education, 2019-2020 Four-Year Cohort Graduation, Cohort Dropout, a-g Completion, and State Seal of Biliteracy Report [PowerPoint slides], <https://www.rcoe.us/home/showpublisheddocument/3802/637581503682670000> (accessed May 2023)
48. Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014) Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs, *Econometrica* 82(6), 2295–2326. <https://doi.org/10.3982/ECTA11757>
49. McCrary, J. (2008), Manipulation of the running variable in the regression discontinuity design: A density test, *J Econom* 142(2), 698–714. <https://doi.org/10.1016/j.jeconom.2007.05.005>
50. Zimmerman, S. D. (2014) The Returns to College Admission for Academically Marginal Students, *J Labor Econ* 32(4), 711–754. <https://doi.org/10.1086/676661>
51. Li, C. H., and Zafar, B. (2023) Ask and You Shall Receive? Gender Differences in Regrades in College, *Am Econ J: Econ Policy* 15(20), 359–394. <https://doi.org/10.1257/pol.20210053>
52. Cataldi, E. F., Bennett, C. T., and Chen, X., First-Generation Students: College Access, Persistence, and Postbachelor's Outcomes, <https://nces.ed.gov/pubs2018/2018421.pdf> (accessed May 2023)
53. California Community Colleges Chancellor's Office, Management Information Systems Data Mart: Transfer Velocity Cohort [Database], https://datamart.cccco.edu/Outcomes/Transfer_Velocity.aspx (accessed October 2023)
54. National Center for Education Statistics. (2022) Undergraduate Retention and Graduation Rates, <https://nces.ed.gov/programs/coe/indicator/ctr/undergrad-retention-graduation> (accessed October 2023)

ABOUT THE STUDENT AUTHOR

Joleen Chiu graduated in June 2023 from the University of California, Los Angeles (UCLA) with a joint M.A. in Economics and B.S. in Mathematics/Economics. Her research interests lie in public economics, particularly in policies that address inequality and access to education.

PRESS SUMMARY

This paper evaluates the college enrollment impacts of guaranteed admission agreements between California State University San Marcos (CSUSM) and its local school districts. The author finds that the admission agreements encouraged students to shift away from enrolling at two-year institutions towards four-year institutions, particularly at CSUSM and at any California State University (CSU). The program also disproportionately affected students from underrepresented backgrounds (e.g., first-generation, nonwhite, and low-income), suggesting that smaller-scale guaranteed admission programs have the potential to increase representation at four-year public institutions and encourage underrepresented students to consider higher-quality postsecondary institutions.

Conflicting Socio-Cultural Attitudes and Community Factors Resulting in Backstreet Abortion in Cato Manor, KwaZulu-Natal

Chloe M. Sachs

School for International Training, Brattleboro, VT

<https://doi.org/10.33697/ajur.2023.095>

Student: chloe.sachs@me.com*, c.m.sachs@wustl.edu

Mentor: john.mcgladdery@sit.edu

ABSTRACT

Abortion in South Africa is a complex topic, rife with augmenting and limiting political, social, religious, and cultural factors. This study explores multiple age groups within Cato Manor and whether the attitudes towards abortion and factors impacting the choice of where and whether to receive an abortion differ. The study employs a general qualitative approach with inserted narratives highlighting individual perspectives and stories. The sample cohort was identified using purposive sampling of women, obtained via convenience through the gatekeeper in the community. The participants are from two generations, and two expert interviews were conducted to gain various perspectives. The sample population is from the greater Cato Manor community, and the interviews were conducted in a semi-structured format. The main goal of this study is to understand how socio-cultural factors impact abortion attitudes within each generation in Cato Manor and if community attitudes impact women's decisions on how and where to obtain medical abortions. The findings show that negative attitudes towards abortion persist in Cato Manor due to religious and cultural rationale. Greater acceptance occurs among younger generations following abortion legalization in South Africa, yet prejudice remains. The research reveals a range of factors that may pressure women to obtain illegal abortion options to avoid community shame.

KEYWORDS

Abortion; Pregnancy Termination; South Africa; KwaZulu-Natal; Cato Manor; Backstreet Abortions; Abortion Attitudes; Illegal Abortion

INTRODUCTION

This study gathered perspectives on how attitudes towards abortion may have shifted over the last few generations in Cato Manor and attempted to determine the main cultural and societal factors influencing choice when considering an abortion. The study was situated in apartheid history, traditional history, and abortion legislation to explore the juxtaposition between government legislation and societal expectations. The aim of this study was to understand the primary considerations and struggles of women looking to terminate their pregnancy medically and why women may choose backstreet abortions when there are legal options in South Africa.

The following four research questions formed the framework guiding this study.

1. How do socio-cultural factors impact abortion attitudes among two generations within the community?
2. What are the reasons women choose to terminate a pregnancy?
3. How do women decide where and how to obtain an abortion?
4. Why do women still seek backstreet abortions when legal options are available in South Africa?

The researcher aimed to understand the role of abortion in South Africa and connect the findings gathered from these four overarching questions. Abortion is a relevant topic specifically in South Africa because it is one of few countries in Africa where abortion has been legalized, within a continent where maternal deaths due to illegal abortion are the highest in the world.¹ However, even after legalization, women are dying at high rates due to backstreet abortions. The Constitution is unparalleled in its abortion liberalism, yet it is not representative of the conservatism that is rampant in the Zulu and Christian ideology of the majority of its citizens. Limited research has been done on why women seek out illegal abortions when safe, legal options are available. As a result, it raises the question of the impact of legalization and why negative attitudes towards abortion remain.

This study was conducted in Cato Manor because the population is primarily black South African and of Zulu origin. It represents many typical Zulu townships in KwaZulu-Natal. It is a peri-urban area with access to the larger city of eThekweni while retaining a

strong sense of community. Families often live in intergenerational housing, and Zulu customs are passed down through families and remain widely practiced. Abortion is a relevant topic within the population, and little research on abortion attitudes has been done among Zulu populations in South African townships. The overarching goal of this study was to contribute to a body of research to understand why rates of unsafe abortion are high in South Africa, as this will provide evidence in the national mission to lower maternal mortality rates.

Context

Pregnancy termination is a morally and politically contentious issue debated in many regions globally. Nations around the world have differing and uncertain laws surrounding abortion. As of September 2022, abortion is prohibited in 23 countries and can be obtained in the first 12 weeks at the woman's request in 76 nations.² The remaining countries fall within these two political extremes. Therefore, there is a tremendous variety regarding the status and experience of abortion in each nation. A further complication is that legally promulgated termination rights are not always offered or easily accessible. Legislation is not the only factor impacting abortion access. Worldwide one in five pregnancies end in abortion.³ This shows the massive scope of abortion worldwide and its relevance in academic health studies. The UN Human Rights Council has designated abortion as a Human Right and designated women's health as central to global health rights.⁴

Despite the geographical proximity, even within the continent of Africa, there is a spectrum of abortion legislation and community opinions. Only three countries, Cape Verde, South Africa, and Tunisia, permit abortion without situational restriction. Abortion is completely banned in 10 countries in Africa.¹ The scope of the unsafe abortion problem in Africa is exemplified by statistics showing that 1.6 million women in the region need medical treatment following unsafe abortions, and abortion deaths are higher in Africa than in any other continent.¹ However, Africa is not a homogenous geographical region, and each nation has different conditions impacting abortion access. Abortion must be more directly studied in individual communities rather than creating generalized assumptions.

The population sampled for this project was from KwaZulu-Natal on the eastern coast of South Africa, encompassing historical Zululand, the home of the ancestral Zulu people. The specific community of interest was greater Cato Manor. Cato Manor, an urban township in the heart of Durban, has a population of 37,622, and 87% were born in KwaZulu-Natal. Cato Manor has a median age of 25, which shows the population is primarily young, and many females are within a fertile age range.⁵ Therefore, abortion is a topic relevant to the population.

The context of modern South Africa must be situated in its history of colonialism and apartheid legislation. However, Eurocentric historians often leave out the rich cultural history that predates European contact. In South Africa, abortion did not originate with Europeans, as many believe. Many herbal remedies were used among indigenous Africans to induce abortion. Zulus depended on an herbal remedy called *uhlunguhlungu* to terminate pregnancies. Abortion has been performed for centuries; however, medically safe abortions were developed and spread around the globe in the 1950s.³

Abortion is politically and religiously charged due to the ambiguous moral question of when and if a fetus has the right to life. Before 1975, abortion was strictly illegal in South Africa. However, the crime of termination of pregnancy was not strictly enforced, primarily because, in Roman-Dutch law, personhood begins at birth.⁶ After colonization and under the apartheid government, white European moral standards took precedence in South Africa, irrespective of the cultural and moral norms of the indigenous population. In 1975 the Abortion and Sterilization Act 1975 was passed under the apartheid government, making abortion acceptable in specific situations, yet it was not equally available. As with most apartheid-era policies, racial and economic lines divided access. Mostly urban-dwelling, wealthy white women had access to legal abortion services. As a result, many black, poorer women turned to dangerous backstreet options.⁶ The historical inequity of abortion access has implications that have carried forward to the modern day. Furthermore, apartheid ideology entered the abortion field by framing legalizing abortion as a form of population control among black populations.³ These types of historical connotations associated with abortion may be a contributing factor to abortion opinions among black South Africans.

The Choice on Termination of Pregnancy Act (CTOPA) was enacted in South Africa in 1996 and inscribed in the Constitution with the rise of a democratic, post-apartheid government. Under 13 weeks, a woman may obtain an abortion under any circumstances, and from 13 to 20 weeks, a pregnancy may be terminated for specific reasons.⁷ Abortion was fully legalized due to a three-pronged argument: The lack of access among poor women, high levels of backstreet abortions and maternal mortality, and the constitutional right to body integrity.⁶ According to Johnston's statistics, a leading expert globally on abortion statistics compiled from various sources, less than 1,600 abortions were legally performed each year before the CTOPA was passed.⁹ After abortion was fully legalized, legal abortions jumped into the thousands. In 2020, 108,301 legal abortions were recorded.⁹

South Africa is one of the few African countries where abortion is legal; however, the rate of illegal abortion remains high. In South Africa 25% of the deaths due to miscarriage result from unsafe abortion practices.¹⁰ Abortion options fall into three sectors: public, private, and informal. While abortions are technically free in South Africa, only one in ten public clinics offer abortion services. 20% of abortions are conducted in the public sector, 26% are completed illegally, and 54% are performed in the private sector. Therefore, the least utilized option to terminate pregnancies is governmentally provided abortion. Why do women seek expensive and illegal abortions when free, legal, and safe options are available? There are likely determinants outside the legal system impacting abortion availability and choice.

Previous studies on abortion attitudes in South Africa

Despite the vast amount of literature on Zulu culture, the studies on abortion and culture are limited. One exception, “Is there room for religious ethics in South African Abortion Law,” states that traditional Africans maintain that a person is made at conception, and abortion should be treated as the murder of a fetus.¹¹ However, Jooee describes abortion in the context of Traditional African Religions, which contributes to but is not equivalent to Zulu culture as it encompasses Zulu beliefs as well as other cultural traditions. Jarvis and Mthiyane (2018) included a short line regarding pregnancy termination summarizing the rituals and cleansing ceremony necessary to prevent negative consequences after an abortion.¹² No further research regarding Zulu practices and attitudes towards abortion exists in current literature within the public domain.

The relationship between religion and abortion has been further studied, as religious sects often have strict views on abortion. Christianity is well researched, as it is the dominant religion within the country and is central to the lives of most South Africans.¹³ According to Jooee,¹¹ 86% of South Africans identify with a Christian sect. The study by Mosley, Anderson, et al. found that more religious South Africans report more negative attitudes towards abortion than less religious ones.¹³ Furthermore, Albertyn explains how the pro-life argument in South Africa has historically been dominated by religious groups condemning women who get abortions as murderers. However, in recent years the pro-life movement has moved away from a singularly religious agenda and instead focuses on the physical and mental risk of abortions to women. Church groups still morally denounce women who get abortions, especially illegal abortions to.¹⁴ The continue to label women who obtain abortions as “murders” of unborn children.¹⁴ Christian attitudes against abortion have remained staunch even after abortion legalization. Mosely, et al. (2017) further confirms that religious individuals are more likely to have negative attitudes toward abortion. They state that in the specific case of ‘poverty leading to abortion,’ Christians and all religious individuals were more likely to display negative perceptions towards pregnancy termination. They were more likely to report that abortion is always wrong independent of context.¹⁵

Much has changed since almost 30 years ago when abortion was legalized after the passage of the CTOP Act. A new generation has grown up never knowing their nation without abortion promised as a constitutional right. Before abortion was legalized, a national survey conducted in 1994 showed that 68% of South Africans opposed abortion legalization, while a similar study in 1995 found that only 21% of the population supported a woman’s autonomous choice.¹⁴ However, after abortion was legalized, recent studies support that the younger generation is more open to accepting abortion. Mosley, King, et al. split their participants into two age groups, 16-45 and 45+. They found that respondents 45+ were more likely to report abortion is “always wrong.” They concluded that there is a difference in abortion attitudes between the two generations.¹⁵ Nilsson, et al. (2020) focused not on abortion but on intergenerational differences regarding sexual expectations set in a rural setting provides some context for intergenerational differences. It states that the older generation usually adheres to traditional beliefs where sexual activity is not practiced until after marriage. However, the younger generation is more modern and open to sexual expression.¹⁶ The study further revealed discomfort among parents discussing sexual matters due to the sensitivity of the topics and how children often turn to their friends for discussions instead of with adults.¹⁶ Lastly, the differences among generations have been viewed as a result of different moral expectations and ideas of sexuality between age groups.¹⁶ Abortion can be a product of premarital sexual activity, and it is likely that the lack of open conversation regarding sexuality among generations includes avoidance discussing about abortion. Ramiyad and Patel (2016) looked at adolescent knowledge and attitudes about abortion. It used a sample population of 150 secondary school students in urban KwaZulu-Natal. Results from this study show that despite LO classes on reproductive health, only 80% of learners knew abortion was legal, and they had even less information regarding the specific terms of the Act. In addition, 20% of the young generation reported that abortion is acceptable. The study concluded that secondary school students have limited knowledge about abortion despite increasing education.¹⁷

Many factors contribute to high abortion rates. One significant factor in South Africa identified in the literature is the high unintended and teen pregnancy rates. Masanabo, et al. (2020) studied the reasons women legally terminate their pregnancies used a sample of pregnant women in a rural area in the province of Gauteng in South Africa. The results of this study show that 24% of participants cited ‘wanting to focus on studies’ and 23% cited ‘not being ready to be a parent’ as the top reasons for seeking abortions among the twelve options. ‘Experiencing financial difficulties followed closely behind.¹⁸ This study occurred in a rural environment, so the data may differ in an urban setting. Frederico, et al. researched ‘Factors Influencing Decision-Making Processes among Young Women’ in Maputo, Mozambique, just across the border from South Africa. They identified four main

factors influencing the abortion decision process. They are “(1) women’s lack of autonomy to make their own decisions regarding the termination of the pregnancy, (2) their general lack of knowledge, (3) the poor availability of local abortion services, and (4) the overpowering influence of providers on the decisions made.”¹⁹ They state that in specific cases others can make the abortion decision for the woman using power and pressure, often the partner. Furthermore, they include familial pressure within this category as well. They identify that safe abortions are only available to fortunate individuals and are not equally distributed amongst the population due to educational opportunities and resources.^{19,20}

The role of men in abortion decisions was studied in low and middle-income countries in a review article published in 2022 by Strong. After aggregating data from 37 sources, the author concluded that men impact women’s abortion choices both directly and indirectly. Strong noted that men were most frequently involved in abortion decisions due to controlling finances and resources. They create the conditions that allow women to access abortions. The father’s reaction also impacted the women’s choices of whether to seek an abortion.²¹ Of these factors, sources disagree on the most important, as all interact in complex mechanisms to impact the often-difficult choice to terminate a pregnancy.

History, social factors, and stigma all contribute to how women choose to obtain an abortion. Although abortions are legal in South Africa, many women still obtain illegal backstreet abortions despite legal risks. According to Chemlal and Russo, women take two paths to get illegal abortions. The first is attempting to obtain a legal abortion and running into difficulties, and the second is just going straight to a street abortion provider.²² The legal and illegal abortion sectors have overlap, they influence and impact each other. Research shows that the main factors that affect a woman’s choice to get an illegal abortion are lack of education about abortion rights, access, community stigma, and mistrust of healthcare workers, which result in numbers of unlawful abortions comparable to before abortion was legalized in 1996.³ Furthermore, Kaswa states in a case study that women’s lack of knowledge about CTOPA legal rights and the community perspective of a negative hospital experience are significant barriers to abortion access. This results in frequent backstreet abortions, especially in rural areas.²³ Meel (2022) in the Transkei region of South Africa within a hospital setting states that confidentiality breaches, staff behavior, and long waiting periods contribute to women choosing backstreet options rather than legal ones. Women are willing to risk complications and death in exchange for privacy and moral treatment.²⁴ Women even receive inhumane treatment by medical staff when attempting to access post-abortion care. Nurses and doctors mocked patients’ pain as something they deserved after having an abortion. Patients delayed their care due to fear of stigma and judgment. Women are even less likely to seek post-abortion care if they previously attempted an illegal abortion, despite being the group most in need of medical intervention.²⁵ Finally, a qualitative study published in 2021 conducted in Cape Town found that 11 out of the 15 women who had abortions used aloe vera, traditional remedies, or purgatives to induce pregnancy termination. Furthermore, four accessed an unlicensed provider and took unprescribed medication. Harries, *et al.* also found that most women turned to friends who had informal/illegal abortions for advice and information about options. It was also stated that women contacted numbers on abortion advertisements to obtain abortions.

Three methods of informal abortions were identified “drinking a mixture prepared by a traditional healer; contacting a non-licensed provider who provided oral abortifacients; and preparing or buying a mixture for a self-managed abortion.” Participants stated their reasonings for seeking informal abortions were privacy and time effectiveness. Furthermore, the study found little research has been done on why women choose illegal abortion options.²⁶ Overall, there is limited information regarding why women prefer backstreet abortions over their legal counterparts, especially in KwaZulu-Natal.

METHODS AND PROCEDURE

The project employed a generic qualitative approach, best utilized for a general study such as this one.²⁷ The researcher conducted semi-structured interviews, which are often used for qualitative data collection,²⁷ to create interview transcripts on the generational attitudes toward abortion and the options to obtain an abortion. Interviews were compiled and analyzed by a single researcher, Chloe Sachs. Thematic analysis was utilized by constructing thematic divisions to develop a research framework before interviewing. The acquired data was analyzed using these preliminary categories, but the categories shifted after the interviews were transcribed and coded.²⁷ The final three themes that emerged were: Socio-cultural Factors, Factors Driving Abortion, and Historical to Modern Backstreet Abortion. Theories were constructed after data was collected, analyzed, and situated within the literature and expert sources. Finally, the research paper used perspectives and quotations from private abortion organizations, community members, and historical data to understand how perceptions of abortions have developed over generations to influence the abortion choices available to women.

Sampling

Convenience, purposive, and snowball sampling, as indicated below, were used to interview community members and ascertain their perspectives on abortion. A community liaison recruited participants within the predetermined age ranges. She thoughtfully recruited participants by informing them of the research topic and by preliminarily confirming their comfort in discussing abortion. After conducting interviews with individuals known to her, the researcher asked them for additional contacts willing to

be interviewed. This was a limitation of the study, as only individuals in the community known the liaison and her participants were included in the research. Lengthy interviews with many participants provided various ideas regarding community views on abortion in Cato Manor. The researcher obtained a spectrum of perspectives and information to analyze. Interviews with experts were also conducted. A representative from Marie Stopes, a private abortion clinic in Durban, shared information on the private legal options for abortion in South Africa and anecdotes about common abortion cases. A second interview was conducted with an expert on family planning. She shared invaluable information about options for women after an unexpected pregnancy and stories of misinformation regarding abortion options among women. Due to the sensitive nature of her work, the second expert asked to be identified solely as an expert in family planning to maintain anonymity.

The researcher initially planned to interview 5-6 members from each of the 18-30, 45-60, and 60+ generations. These age groups were chosen, respectively, to represent a generation after abortion was legalized in 1997, a generation for whom abortion was legalized during their childbearing years, and women who experienced pregnancy before abortion was legalized.⁸ However, once the interview process started the liaison shared the opinion that interviewing two generations would be more possible within the time constraints. Therefore, the investigator elected to change the sample population to two generations, women ages 20-30 and 50+.

Because the research method was qualitative, the researcher focused on in-depth interviews with a limited number of individuals rather than gathering a magnitude of data. The project focused on understanding the factors influencing abortion options and the socio-cultural pressures facing women when terminating a pregnancy. The answers to these questions differed between individuals, but 15-16 interviews from each generation were sufficient to identify patterns and themes. The aim was 30 total interviews to add further validity to the research. Because many themes were studied, the total number of participants was raised to reach data saturation. The final sample population included 31 women, 16 members of the younger generation, and 15 older than 50. The youngest participant was 20, while the oldest was 73. The mean age of participants in the younger generation was 24.3 years, while the mean age of participants in the older generation was 60.3 years.

A systematic review study conducted in 2022 states that data saturation occurs around 12-13 interviews.²⁸ 15-16 conversations within each age group were sufficient for data saturation. Data saturation was reached, as during the final ten interviews, no new information or themes were generated beyond additional personal stories. The ratio of different opinions and frequency of themes referenced remained consistent after the first 20 interviews.

The data collection site was Cato Manor, specifically KwaMasxha and Chesterville, and only Zulu women were included. There were no criteria regarding religion or disability. English-speaking participants were preferred; however, Zulu-speaking members were not excluded. Nine interviews among the 50+ participants were translated or partially translated by the community liaison. No conversations in the 20-30 generation required translation. Interviewees were identified through the community liaison and contacted by the liaison in-person and via WhatsApp.

Data collection instruments and methods

Data was collected through one-on-one semi-structured interviews with participants. An interview guide with seven topics was used to initiate a dialogue between the participant and researcher on abortion. The first five topics were conversational, and the final two were direct questions because they were less personal. The conversational prompts allowed the participant to avoid topics they were uncomfortable with, and the interviewee led to the hoped-for conversation without direct questioning by the researcher. Many prompt questions were pre-written to allow for flexibility. The questions were not asked in order, and some were skipped depending on the direction of the conversation. Many of the questions were not asked during the interviews because they were addressed without prompting. The researcher let the participant control the interview's direction, as Anderson and Kirkpatrick suggest.²⁹ The interviewer's primary role was to keep the conversation flowing and allow the participant to speak about any topics they wanted about abortion. Some interviews were prolonged due to time spent discussing non-abortion-related issues. The regular flow of conversation established trust between the interviewer and participant. However, the conversational style prevented any participant from feeling interrogated or forced to answer questions that made them uncomfortable. Additionally, the participants shared more information in a semi-informal conversational structure rather than a structured interrogatory. The interviews were conducted at participants' homes in KwaMasxha and Chesterville, except for two interviews at Coweys Corner due to inclement weather. Interviews ranged from 15 minutes to 1.5 hours. The community liaison organized transport and the School for International Training provided funds to cover the cost of transportation and to compensate interviewees for their time.

Since abortion is a sensitive topic among some participants, before beginning the interview, each participant was asked, in their preferred language, if they were comfortable discussing abortion. The researcher made every effort to prevent significant discomfort from being evoked in the participant during a discussion on abortion. A conversation was had between the interviewer

and participant to determine the interviewee's comfort in discussing abortion. If discomfort was expressed or suspected, a general alternative interview was offered to maintain confidentiality. The alternate interview consisted of general questions about Cato Manor and the individual's experience living in the community. Throughout the interview, the interviewer remained aware of body language and cues of discomfort, and verbal checks on comfort in continuing the discussion were conducted at regular intervals.

Data analysis

The researcher recorded all conversations and responses to draw out themes and compare respondents' views with those noted in the existing literature. Current perspectives were juxtaposed against older participants' historical views and recorded academic histories. Where appropriate, short narratives that highlight issues were inserted into the analysis. The researcher's main objective was to maintain the integrity and structure of participants' narratives regarding abortion. It was essential to avoid misrepresenting the interviewees and preserve the stories' personalities during analysis.²⁹ The study was created by situating the quotes and perspectives of participants within the framework of their generation and South African history. Attitudes and biases towards abortion were connected back to historical and generational causes when possible. Specific themes were chosen as a framework to contain and analyze community perspectives without changing important wording or meaning.

The interviews were recorded via Voice Memo and Otter AI, an automated transcribing online platform, with participants' verbal consent. All audio recordings were stored on a password-protected device and were deleted immediately after the project concluded. Stories and quotes were directly transcribed based on relevance to abortion. Sections pertaining to topics other than abortion were not transcribed. The researcher then coded the data by counting the number of times specific phrases or ideas were mentioned and pulled useful quotes. The coded topics were 'abortion is a sin,' 'I don't judge,' 'Zulu rituals,' 'bad luck,' 'young girls are drunk and reckless,' 'consequences of unsafe sex,' 'forced by baby daddies,' 'a child if a gift from God,' 'abortion shouldn't be legal,' 'it's an embarrassment/disgrace,' 'parents never mentioned abortion,' 'learned about abortion in Life Orientation,' and 'killer/murderer.' Coding was then used to consolidate the final themes from the preliminary framework. Participants only met coding criteria if they directly stated the code topic and did not contradict their initial opinion later in the conversation. When presenting findings, direct quotations were intermixed with the interview response summary. Narratives were purposely included to remain faithful to the nuanced information provided by the informants. Stories were kept in their original wording and format to allow the participants' voices to shine through.

Thematic analysis was used to address the project's question and aims. Themes were developed top-down before the commencement of interviews using current literature but were flexible to change, and new themes emerged after interviews were conducted.²⁹ Analysis was completed by extrapolating a more significant meaning from the findings, combining the results of multiple themes, and connecting primary data back to the literature. Literature was purposefully only sought out from low and middle-income countries, mainly within South Africa, to compare and contrast with similar contexts. Three final themes were decided upon, the first with three major subheadings. The themes are organized to show how religion and culture result in negative attitudes towards abortion, and despite education and acceptance among some youth, moral standards take precedence. As a result, the range of social issues that pressure women to get abortions force them to turn to illegal options to avoid community judgment.

Limitations

Limitations of this study included age range. It only constituted women from 20-30 and 50+, so this research did not consider women under 20 or between 30-50. The participants were sampled using convenience, snowball, and purposeful sampling. Only individuals known to the liaison or her acquaintances were included in the study. Therefore, the sample population is not entirely random and may not accurately represent the Cato Manor population. Furthermore, despite the researcher's best intention to separate participants during interviews, some interviews were held within the hearing range of other participants. There may be some bias among participants who overheard others' answers before the interview. Another study limitation may be loss of accuracy and wording during the translation process that was required for some of the older generation participants. The translator did not provide a perfect word for word translation, despite her best attempts. The major limitation of this study was that no participants revealed firsthand experiences seeking abortion services or facing community judgment following an abortion. This may have been due to fear of reprisal. As a result, the researcher focused on secondhand narratives and beliefs among individuals regarding pregnancy termination.

Ethics

Carefully designed, this research ensured that all members were not harmed. Discussing abortion may be a vulnerable topic for some individuals who have personal experiences with abortion. Therefore, before each interview, the researcher asked if the participant was comfortable discussing abortion. If not, an alternate interview on a different topic was offered. In addition, participants were invited to discuss personal issues that may cause discomfort through a conversational approach rather than

questions that directly elicit responses. The participants were invited into conversations through a copy of an illegal abortion advertisement as a prompt. Less personal questions were asked through direct questions.

When starting each interview, it was made clear that the questions regarding abortion did not need to be asked or answered. Individual questions could be skipped and the interview could be called off at any point. Interviewees were allowed to redact or refuse publication of any provided information. In addition, all participants received total financial compensation even if the interview was unfinished.

There was a small financial incentive, and participants were invited to ask the same interview questions of the researcher to facilitate equivalent cultural exchange. In addition, the study aimed to give individuals some information about where abortion stigma originates and knowledge about safe and legal abortion. Some information was shared with participants after the interview about the dangers of illegal abortions for the health of the mother. For example, Participant 18 believed the adverts offered the safest possible abortions. Following the interview, the interviewer shared research about unsafe abortion and that the adverts are usually scams by illegal abortionists.

Since all participants were asked before the interview whether they were comfortable discussing abortion and would be free to stop the discussion at any point, there was no risk of harm to the participants. The interviewer did regular verbal and observational checks to confirm participant comfort. Participant 9 consented to discuss abortion and confirmed she felt comfortable. However, throughout the interview, she gave one-word answers, and her body language was tense. As a result, despite a second follow-up when she again said she wanted to continue, the researcher skipped the final few topics of discussion and instead asked non-abortion related questions about childrearing and family dynamics. The interview was continued to avoid embarrassment by pointing out her discomfort, while moving the topic away from abortion. The participant never verbally stated she was uncomfortable, but the interviewer wanted to err on the side of ensuring that no emotional harm or discomfort resulted from the study.

Oral consent was obtained through a verbal explanation of the interview process, participant rights, and study details. Physical forms requiring written signatures were signed by each participant as well. For non-English speakers oral and written consent was obtained via direct translation by the community liaison. Elderly populations, the 50+ age group, can be characterized as vulnerable; however, the research design ensured all participants fully understood and had all the information to consent in their preferred language.

All informants were asked not to share what they did not want others to know during the introduction. It was made clear that any information could be redacted from the right to publication. The participants were informed they could withdraw consent even after the interview concluded. Furthermore, the participants received an immediate verbal summary during the interview, and the ISP is available on the SIT website and accessible to anyone with internet access.

All participants are anonymous. Since personal information about abortion and shared opinions on abortion were collected, the participants must remain anonymous. Therefore, no names were used or collected, and no unique identifiers such as family names, addresses, physical identifiers, or defining characteristics are included.

All data is on password-protected devices, and voice recordings were deleted on June 1st, 2023. If data is used in the future, the individuals will be recontacted for renewed consent. The data will be available online, but only information consented to by participants.

RESULTS

Socio-Cultural Factors

This section compiles the critical data and quotations found in this study relating to socio-cultural factors. After presentation of the major findings, analysis is organized into three sections. The first section is ‘Get rid of the bad luck’ focuses on Zulu cultural norms and looks at historical Zulu culture and the modern implications on abortion attitudes today. The cultural norms surrounding virginity and womanhood are combined with primary accounts of contemporary cultural practices surrounding abortion. The second subheading ‘Is talking about abortion condoning it?’ focuses on the impact of Christian beliefs, and religion is analyzed within its historical context in South Africa and among Zulu populations. Additionally, the intersection between culture and Christianity is closely looked at. The final subtheme ‘The promiscuous youngsters’ examines intergenerational differences and observes and analyzes the differences and similarities between the two age groups.

Throughout the interview process, the topic of Zulu culture’s impact on abortion attitudes was passionately discussed by participants. All participants were black South Africans of Zulu origin, as indicated by the vast majority of Cato Manor being of Zulu descent, and participants mentioned their culture as Zulu throughout the interview process. A majority of the participants

noted that abortion is unacceptable in Zulu culture. Participant 17 explained the lack of acceptance for abortion in their culture by stating, "In the Zulu culture, we believe the fetus then is regarded as a spirit already, and already a child. In our culture, it already has a spirit."

Additionally, participants provided detailed knowledge regarding the cultural ceremonies required to remove the bad luck after an abortion. More participants may have known more about Zulu abortion traditions but did not share them during the interview because no question was asked to describe the rituals in detail. A 20-year-old woman stated, "I'm a black person. We do traditional things. Yeah. So after abortion, they say you must go to cleanse and do *iladi*, Like a small ceremony for the baby. You can't just do abortion and stay. You going to be followed by ghosts." Participants described *iladi* as having a birthday party for the unborn child, including slaughtering animals, buying presents, and naming the child. Then the woman must apologize to the ancestors and ask for forgiveness for the bad luck to dissipate. Participant 17 further expanded on the impact of abortion within the family. She shared, "It then becomes a generational curse. When your kids have grown, and they want to have kids, and they can't have kids. And when they look into it, it's because you had a couple of abortions." Abortion imbibes bad luck upon the woman, her family, and even the man with whom she conceived the child.

Furthermore, stories are shared within the community by parents and friends about the negative consequences of an abortion. Participant 4 explained that she learned about the effects of abortions in Zulu culture from a story passed down by her mother. *"My mother used to tell this story over and over again. I don't know if it's true or not. But she used to tell us, her friend who went and assisted her friend for abortion. And when she got married she didn't have kids...they were asking for help going to hospital, and they said there's nothing wrong with you. You're fertile, you can have kids, but she was not conceiving. Then the sangoma told her. She had to go back to her mother and said Ma, we must apologize to the ancestors. So they slaughtered a goat, apologized, did the rituals. And then she fell pregnant."* A similar story was shared by Participant 14 of a friend with a sick child who lost her job. By consulting with a traditional healer, Participant 14's friend learned that a past abortion was the cause of her misfortunes. While participants questioned their validity, these folk tales are regularly passed between generations and friends.

Eight individuals directly stated that they identified as Christian, despite no question directly asking about their religious affiliation. Five were older than 50, while three were in the 20-30 age group. No participant verbalized any association with a religion besides Christianity. Additional participants referenced Christianity but did not unequivocally state that they identified with a religious denomination. Five individuals said abortion is a sin. Three were older than 50, and two were within the 20-30 age range.

Religious participants made statements such as, "Christianity tells us not to kill. You know the 10 Commandments? It's one of those things." A powerful statement from a translated interview depicting the conviction of religiosity was, "From her perspective and her beliefs, she thinks abortion is a bad thing because God gives life and who are you to take life. God Almighty is everything to her. Whether it's legal in your country or any other country, she still doesn't approve of it." Participants, especially within the older age group, shared sentiments that even if abortion is legal, it does not mean it should be allowed due to their religious value system. Three participants stated that because they were Christian, they knew nothing about abortion and found it a distasteful subject to discuss. Participant 31 said, "I'm a Christian. I don't know nothing about this." These participants were willing to share how abortion was wrong, but the conversations with Christian individuals rarely led to additional information outside the religious sphere.

Finally, when discussing the main factors impacting abortion, participants addressed the overlap between Christianity and Zulu culture. Five participants used Christianity/religiously/spiritually in conjunction with culture/tradition to describe what caused the community to denounce abortion. No participants mentioned additional factors that led to judgment on abortion besides, "It is something that is unacceptable. I just don't know how to explain it." No participants discussed the linkage between religion and culture, yet many referenced the two together to explain why abortion is wrong.

The participants were split into two age categories, 20-30 and above 50. Sixteen participants were within the 20-30 age range (52%), and fifteen were over 50 (48%), resulting in a total sample size of 31. A higher number of the 50+ participants portrayed anti-abortion attitudes. Seven participants stated that abortion is a disgrace or embarrassment for the woman having the procedure or for her family. Three were young and four were older. Furthermore, eight participants shared that young women are drunk and reckless or girls are to blame and must accept the consequences of unsafe sex. Only two of these eight were within the 20-30 range (25%). All others were over 50 (75%). Blaming young women usually came from the older generation. A 59-year-old woman explained, "It's because the younger people drink a lot, and they sleep around, and they don't want to face the consequences of their actions. That's why they promote it or like it." To some older women, abortion was only a problem of the reckless younger generation.

The younger generation was more understanding of why women need abortions. Interviewees in the 20-30 age range shared that they would not judge if a woman had an abortion. Participant 11 shared that her reason to refrain from judgment was, "Maybe someday it's going to be me doing an abortion after judging someone." Younger participants showed a higher level of acceptance for women who must get an abortion and the factors forcing them towards that decision.

The most valuable story showing a shift within the younger generation regarding abortion and a capacity for forgiveness came from Participant 14. She offered a rare perspective on abortion situations. She was a baby whose mother tried to abort her then proceeded to commit suicide.

"So wherever she left me, she thought I was dead, and then she went to commit suicide... So I would say that abortion has a reason. A person would just not do it because they want to do it. My mom did it because at that time, she was struggling. She already had a kid. She was pregnant with me. So for her, I thought that it was gonna be hard raising two kids, and she was unemployed. She had two baby daddies. It wasn't easy for her"

"Instead, I forgave her. For her to be okay eventually, for me to be okay as well. Because I feel like she wouldn't kill me because I'm her daughter. I don't think a mother would kill. She just ran out of options and choices... Until you wear the same shoes, and then you reason down and introspect why the person does this. You understand not to judge."

The voice of acceptance for women who get abortions were mainly from the 20-30 age range. Some women 50+ showed some degree of understanding towards women who get abortions but were more contemptuous.

Another significant intergenerational difference was in how information was disseminated. Three individuals in the 20-30 generation stated that their parents discussed abortion with them, while elderly participants often laughed when asked if their adults mentioned abortion to them as children. No participants 50+ recalled ever discussing abortion with their parents. They learned about abortion through peers or older members of their community. One young woman's mother spoke to her about abortion growing up and said, "She always tells me whenever it happens that I get pregnant, I must talk to her and not to do an abortion behind her back. Because I'll be faced with problems." Furthermore, two participants, aged 20-30, mentioned without prompting that they plan to talk to their children about abortion. One interviewee shared, "I would. I would. Now, to be honest, I'll speak to my child, especially if it's a girl. I'll speak about everything."

The younger generation also learned about abortion in school. Six participants shared that abortion was mentioned in their Life Orientation classes in high school. All of these participants were in the 20-30 age group. When describing Life Orientation, LO class, Participant 6 stated, "At high school, we were more educated about it because that's where we had certain classes like life orientation, LO class, where they speak about abuse, HIV, and abortion." Despite the word abortion mentioned in class, the older generation of teachers were perceived as giving little information to students. Another participant said LO teachers explained abortion as, "We're just teaching you something because it's in the book. But it's not something that you can go out there and practice." Four participants shared that many teachers were against abortion and shared this perspective with the class rather than providing an objective description of family planning options. No members of the older generation mentioned learning about abortion in school. The sources of information about abortion differed between ages, with younger generations learning from school and parents and older generations from peers and external sources. This suggests that the older women learned from the silence of their mothers and wanted to open a dialogue with their daughters.

"Get rid of the bad luck"

The findings showed the influence of Zulu culture on abortion went beyond what exists in the limited literature. During the interviews, the participants were eager to discuss Zulu culture. Almost all participants mentioned Zulu culture as a significant factor influencing abortion attitudes. Abortion is acknowledged within Zulu culture to be wholly unacceptable and equivalent to murder. Culturally a fetus is a child from conception, creating a decidedly hostile attitude towards abortion within the traditional community. Jorgee backs up these statements reported by participants in his account of personhood originating from birth in Zulu traditional religions and culture.¹¹ Zulu ideology starkly contrasts with the Afrikaner perspective on abortion, in which personhood begins at birth.⁶ White colonialist moral and legal standards marginalized Zulu cultural norms for centuries. As such, the firm adherence of Zulu South Africans to their traditional beliefs regarding abortion may be a way of resisting their oppressors. Zulu individuals may be less open to accepting abortion, as it concedes to the morals of their prior colonizers and oppressors. While culture was not cited as the sole factor impacting abortion attitudes, it was a significant contributor.

Brief comments by Jarvis and Mthiyane state that a cleansing ceremony is traditionally performed following an abortion to "get rid of bad luck."¹² However, the current study gained more insight into the specific components of an *iladi* cleansing ceremony. The process includes having a birthday party for the child, apologizing to the ancestors, naming the child, and slaughtering an animal.

“And because you are a Zulu girl there are Zulu customs that need to be followed if you have a child. So this child is going to be a blockage for you when you grow up. Because now you have to do a ceremony for this child, you need to bathe and take out the bad luck.”

Despite Zulu culture condemning abortion, it is acknowledged and discussed, and the specifically defined rituals show that abortion has been present in Zulu communities for a long time. The study found that the fear of bad luck in Zulu culture significantly leads to community judgment about abortion. Abortion is characterized as an affront to the ancestors who will bring bad luck to the woman, her family, and her partner. The transmission of bad luck to individuals connected to the woman is not discussed in current literature. Zulus believe that bad luck from angering the ancestors prevents individuals from being successful. If the baby is not acknowledged, it will haunt the woman. The cultural concept of bad luck is essential to understanding abortion in a Zulu context.

There is a clear overlap between modern medicine and traditional healers when considering abortion. Women have medical procedures to obtain abortions, which are legal and accepted within South Africa, then consult with a *sangoma* who blames their misfortune on the past abortion. There is discord between the legal status of abortion as a safe medical procedure and the negative spiritual results of abortion in Zulu culture. While medically, there may be no complications from a legal abortion, spiritually, Zulu women fear for the well-being of themselves and their families. For Zulu women, unlike women of other cultures, there are additional cultural considerations when getting an abortion. Historical context in South Africa states that Zulus used *ublungughlungu* to terminate pregnancies;³ however, no women within this sample population mentioned using traditional medicines or going to a *sangoma* for an abortion.

Finally, stories passed within the community impact community perceptions of abortion. The story passed down to Participant 4 by her mother exemplifies intergenerational passage of abortion perceptions as cautionary tales. Generations of colonialism and apartheid have oppressed Zulu culture, and as a result, storytelling is a meaningful way of passing culture between generations. Rather than strictly telling her daughter that abortions will lead to bad luck and an inability to produce future children, stories were used as a powerful tool to share cultural expectations and messages for Participant 4. Furthermore, the story from Participant 14 of a friend who attributed her hardships to her abortion after consulting with a *sangoma* proves how stories also pass within age groups. Negative cultural attitudes surrounding abortion are made credible by personal accounts of a ‘friend.’ Culture is a powerful tool of community control. In the case of abortion, the cultural assertion of bad luck compounded by anecdotes within the community has led to a widespread belief that abortion results in misfortune and disgrace.

“Is talking about abortion condoning it?”

The data demonstrates that Cato Manor residents considered religious identity, specifically Christianity, when discussing abortion. The importance of religion is directly supported by Mosley, et al., who state that Christianity is a pillar in the lives of most South Africans.¹⁵ As a result, faith plays a role in women’s decisions about whether or not to get abortions. Religious participants in this study shared strong perspectives that abortion is undeniably wrong, no matter its legal status. Data from this research suggests that the Constitution is out of touch with the nation’s majority religion regarding pregnancy termination legislation. As such, women will continue to die in record numbers due to illegal abortions because they will reject the safe practices available to them, as a consequence of Zulu and Christian social stigma related to abortion. Jooze bolsters this perspective by stating that although South Africa’s constitution is liberal, this often does not represent the sentiments of its people.¹¹ Furthermore, this study shows that many Christians view abortion as murder or are not willing to discuss abortion. Abortion is a taboo subject among religious individuals, a topic so incompatible with their value system it cannot be talked about. Having an open conversation about abortion can be equated to condoning it. Among the sample population, avoiding discussions on abortion was the best way of displaying distaste for the practice among religious individuals.

The older generation was slightly more likely to bring up religion more often in combination with abortion than the younger generation, and is suggestive of the Christian reaction to abortion has shifted over time. As stated by Albertyn, the pro-life argument has changed within South Africa. Religious groups are no longer leading the campaign against abortion, and the momentum has changed to physical and mental concerns for women. However, the Christian agenda remains a significant pillar of pro-life movements, albeit with less influence.¹⁴ In this study, young participants were less likely to use religion to explain why abortion was wrong than older participants, validating Albertyn’s argument.

Finally, research often disaggregates data by religious and cultural groups. However, identity must be considered from the perspective of intersectionality. While this pioneer of antiracism focused on race and sex, in this study intersectionality brings into contact race, sex, religion, culture, and age. This sample population comprised a range of individuals with unique and overlapping aspects of their identity. Abortion judgment was frequently described as stemming from both religion and culture. For some participants, it was not just one moral standard that contributed to negative perceptions of termination but the combination of

Christianity and Zulu norms. For others, the reasons abortion was wrong were unexplainable, not a consequence of culture or religion, nor definable to any other source. The reason abortion judgment is challenging to identify is because of the range of cultural, religious, and social factors that contribute to one's moral compass and ideas of personhood. It is not easy to parse out these individual values within the time of short interviews.

"The promiscuous youngsters"

The two age groups within this study represent women raised before abortion legalization and women growing up in an environment where abortion has always been an unalienable right. Participants displayed a range of opinions on abortion across age groups. As found in the literature, 68% of the population opposed abortion before legalization,¹⁴ and this percentage has changed minimally even after the legalization. An almost equal number of participants stated that abortion is an embarrassment or disgrace across age groups. Both young and older women displayed internalized bias against abortion as a shameful action. As seen in the Zulu Cultural Norms and Christianity data among the sample population, cultural and religious beliefs still impact abortion beliefs today. As long as negative cultural and religious attitudes towards abortion are passed down intergenerationally, shame will not decrease.

Most women blaming abortions on young, reckless girls were over 50. Rather than seeking to understand the complexity of reasons women obtain abortions, the older generation saw it as an easy out for careless and promiscuous youngsters. Abortion was legalized after the older generation was of childbearing age. Therefore, there may be a degree of resentment that the modern era is utilizing government-sponsored abortion to solve unplanned pregnancies when older women never had the option. During apartheid, abortion was a privilege for only the white and wealthy.⁷ The older generation likely still has internalized bias towards the practice.

The younger generation seemed better at understanding their generation's and peers' plight and was more likely to avoid judgment. Instead of trivializing abortion as something that happens among careless girls, they had a better perspective on the many reasons women may seek an abortion. The literature supports this argument as in a similar study, respondents over 45 were more likely to state that abortion was 'always wrong.'¹² Younger generations are more likely to understand the nuance of reasons women get abortions. The story shared by Participant 14 best highlights the ways change has occurred. This was a woman who, despite her mother's attempt to abort, forgave her mother and tried to understand the difficulties she was facing. She understood how hard it must have been in her mother's shoes. The forces of poverty, multiple children, and a lack of a support system were struggles she understood and could empathize with. No similar willingness to understand why women get abortions was found among the older generation.

After determining that attitudes towards abortion differed among the two generations, the question is why? Cultural and religious attitudes have remained constant, so what has changed?

The major shift was abortion education in the home and school. Older women stated that their mothers never discussed abortion with them, it was even laughable to imagine conversing sexual activity in the home, but some said they discussed abortion with their own daughters. Furthermore, members of the younger generation explained that their mothers spoke about abortion and that they would teach their future children. Over the last few generations, a significant shift has occurred. This is further explained by studies showing that older women retain traditional views of sexuality and morality and, therefore, would be uncomfortable discussing abortion with their children.¹⁶ Culturally and religiously Zulu and Christian women are supposed to maintain their virginity until marriage. However, this current study and others show some of these values have lessened among the younger generation in favor of openly discussing sexual relations. From this data, it can be concluded that the younger generation, while maintaining their adherence to faith and culture, has been influenced by biomedical sexual education in Life Orientation classes. Young women know they are more educated than their mothers about abortion because they learn about it at school and in their home.

However, not all younger generation members were less judgmental and more empathetic about abortion. Many made derogatory and condemnatory statements toward women who get abortions. The method in which abortion is taught in schools gives a clue as to why. This study's primary data and Ramiyad and Patel's study show that although abortion is part of the LO curriculum, the information disseminated about abortion is often biased. Participants shared that LO teachers were often Zulu and held negative views of abortion which they shared with their class. They made it clear they were teaching abortion because they were forced to, not because they condoned the practice. As a result, earlier studies show that only 80% of learners knew abortion was legal, and 20% believed it was acceptable.¹⁷ School-age students are in the formative years of their moral code and belief system. If teachers are supposed to provide unbiased and accurate information and inject personal opinions into abortion education, it will likely affect children's perspectives.

The increase in abortion education in schools has caused intergenerational changes in the perception of abortion, and the younger generation is more open to abortion. However, retained stigma among mothers and teachers has been passed down to learners and has limited intergenerational changes in abortion attitudes.

Historical to Modern Backstreet Abortions

The following section, 'Historical to modern backstreet abortion,' used participant stories to explain the progression of backstreet abortions and the role of illegal abortion among modern-day women. It uses historical documents and past studies along with modern day narratives to understand the changing role of abortion in South Africa and specifically Cato Manor.

Each interview began by showing the participant an image of posters advertising illegal abortions found regularly on stoplights and walls around eThekweni, the central city neighboring Cato Manor. 29 participants (94%), all but two, confirmed that they had seen the posters before. Two participants confirmed that they began seeing the signs in the early 2000s after abortion had been legalized. However, illegal or 'backdoor' abortions occurred far earlier than when the CTOPA was passed. A participant in their late 50s explained, "Abortions are very old. They were done before. Because before, it wasn't legal; that's how the Mamas started abortions. That's how the backdoor started about. Since then, they're popular and still do it." According to interview sources, there have been women in the community who have been abortionists for as long as they can remember. Three stories were shared of illegal abortionists who worked and advertised their services within Cato Manor. A 58-year-old woman stated that when she was 16, "My neighbor, she was doing abortions. She was a family friend." further shared, "They were well known in this township. They were well known. People come from Umlazi and other townships to Chesterville... One passed away in town because she came here. One had the process done in Chesterville and got in a taxi. But as she was approaching the taxi rank, when she got off the taxi, that's when the process started aborting. And she was found in the toilet with the baby." Before abortion was legal, women got illegal abortions in Chesterville, a subarea of Cato Manor. This story also depicts that women doing abortions were generally known within the community and well utilized.

In addition, an understanding of the dangers of illegal abortion was gained during interviews. Eight women mentioned that backstreet abortions are dangerous, and women died, especially before legalization. The expert from Marie Stopes shared an experience of a woman who was cajoled by an illegal abortionist in the lobby of Marie Stopes. The patient was waiting for her appointment when a man approached her and offered her the same procedure for a lower price at an 'abortion clinic' next door. She followed the man and was brought into an unsanitary room where she was told to squat on a dirty mattress. She quickly realized she had been bought into a backstreet abortion scam and returned to Marie Stopes. Furthermore, the expert shared that many women come to Marie Stopes after a failed illegal abortion. This results in excessive bleeding during the procedure and complications that can send a woman to the hospital and endanger her life. Illegal abortionists are easily accessible and spread out through the Cato Manor community and the city of eThekweni, but provide unsafe and unsanitary services.

Illegal abortions are still sought in Cato Manor, despite legal procedures offered at public hospitals and private clinics. Secrecy and cost were the main reasons brought up by participants to explain why women would choose an unsafe abortion option. Private legal clinics are outside of most women's price range, and one participant with knowledge about the private sector stated, "they are costing more than R2,000 for a procedure." In contrast, the posters in town offer abortions for 'cheap,' which appeals to many women desperate to get an abortion but without the funds. The expert from Marie Stopes explained that many women come to the clinic without the required funds for a procedure, and while the clinic can occasionally subsidize operations, they often have to turn women away.

Furthermore, other participants mentioned that public hospitals, the only free abortion option, have long waiting periods, and clinical staff often mistreat women. Women can visit public clinics up to three times without getting a procedure and panic as their pregnancy progresses. Participants stated that women choose backstreet options over hospitals due to mistreatment or a lack of privacy. An older participant shared, "It's an embarrassment; it's a shame. In our hospitals, if you have a sickness, it's not private. 'These ones are going to have an abortion on this side, the ones who have HIV.' Why would I go to a hospital if this is how I would be treated... The way these nurses treat us, I'd rather go to these people." This quote shows the lack of privacy afforded to women in clinics and the fear of being outed for having an abortion. The 66-year-old woman stated she would instead go to a backstreet abortionist and risk her life than be subject to the humiliation of going to a hospital and being mistreated by the staff.

Abortion is legal in South Africa, yet almost every participant was intimately familiar with illegal abortion options in eThekweni. Most participants had seen the illegal abortion advertisements and had a friend or heard a story of someone who had obtained a backstreet abortion. Backdoor abortions are a remnant of pre-legalization in South Africa, yet their popularity endures. Participant 17 confirmed the existence of backstreet abortionists in Cato Manor for at least the last 40 years, as she was 16 when she learned her neighbor was a backstreet abortionist and was 58 during the interview. Furthermore, participants validated the 1996 justification for abortion legalization by sharing horrific stories of women dying from illegal abortions. Women were found in

toilets at the taxi rank or placed into the ground as the illegal abortionist cried over the coffin. The question remains of why backstreet abortions are still widespread post-legalization.

“What if I don’t have R2000?”

The number of modern stories shared about backdoor abortions corroborates the data claiming that 26% of abortions in South Africa are still performed illegally.¹⁰ Chemlal and Russo state there are two paths that women take to illegal abortions, yet this study’s research differed from this study by only revealing one path.²² Within Cato Manor, from the interview data, women do not go to the hospital first. Rather it is the opposite. At Marie Stopes, patients come in after an illegal abortion failed to terminate their pregnancy. Throughout the research, no participant shared a story of a woman first attempting a legal abortion, then turning to a backdoor option. The only exception was the story of the woman at Marie Stopes tricked by an illegal abortionist, and she returned in horror to the certified clinic. However, in this study, those seeking an abortion immediately choose whether to use a backstreet or legitimate provider.

Much of the choice comes down to cost and privacy. When choosing between public, private, and illegal abortions, women must decide whether to give up privacy, money, or safety. Some women risk their lives for privacy and a low price. The reason for this is the embarrassment associated with abortion created by the cultural and religious shame previously analyzed. Young girls often don’t have access to the R2000 necessary for a private abortion, and privacy protects them from familial judgment. Though not well supported by the literature, other findings contextualize the need for a cheap and secret abortion.

Three previous studies on reasons women get illegal abortions overlap with the theme of hospital mistreatment. Hodes, Kaswa, and Meel agree that mistrust of healthcare workers and negative hospital experiences result in women avoiding government-sponsored abortions.³ The Cato Manor research demonstrates that hospitals, in general, are mistrusted. Privacy is nonexistent, and ill treatment is rampant. Therefore, when getting an abortion, internalized mistrust of the public healthcare system prevents women from attending the hospital. Obtaining an abortion is already shameful in Zulu culture and among Christian individuals. Fear of shame at the hospital adds additional embarrassment to an already traumatic experience. As a result, women would rather go to a backstreet option.

An impressive study in Cape Town was conducted by Harries, *et al.* among women who received illegal abortions. Although the Cato Manor study did not have access to participants who had personal experiences of getting abortions, the second-hand stories in this study resulted in very similar findings. Participants turned to friends for advice, corroborated in this study by children not discussing abortion with parental figures, to find illegal abortionists or other abortion remedies. Similar to this Cato Manor study, women found names of illegal abortionists in advertisements. One difference was women in the Cape Town study reported three ways of obtaining abortions: traditional medicines, illegal abortionists, and self-prepared mixtures.²⁶ While illegal abortionists were discussed frequently and self-prepared mixtures brought up on occasion, no women mentioned traditional medicines. This may be due to the different demographics or the small sample sizes of both studies. The women of Harries, *et al.*’s study were mostly involved in transactional sex and had all had an illegal abortion. In contrast, this study in Cato Manor compiled community views, which resulted in differences in abortion knowledge. Furthermore, the communities were in entirely different regions of South Africa, and Cato Manor is a peri-urban township while Cape Town is an urban city center. Therefore, the Cato Manor participants’ experiences versus those from the Cape Town study unsurprisingly vary, as women likely seek different sources for abortion depending on their geography and community. However, both studies agree that women turn to illegal abortions frequently and for various complex reasons.

Community Factors

The final theme investigates the prominent factors driving abortion within the township of Cato Manor. It brings together the forces leading women to get abortions from primary sources and links them to other studies and societal explanations. The three main factors driving abortions cited by the participants were teenage and school-age pregnancy, the lack of an adequate support system, especially from male partners, and fear of a negative familial response. Other causes were mentioned, but these three were the most repeated throughout the interview, regardless of generation or personal beliefs about abortion. Eleven participants stated those getting abortions are young, teenagers, or still in school. Older participants said, “It’s a dumb teenage thing. You can always prevent yourself from getting pregnant. She doesn’t understand why kids do it.” Others stated that schooling plays a role stating, “It’s mostly students. Most of the time. They are youngsters. They are careless. They think, ‘What are my parents going to say, they sent me to school, and I’m pregnant now.’” Younger participants also stated that teen pregnancy leads to abortion, and 4 out of the 11 interviewees stating teenage pregnancy leads to abortion were in the 20-30 age group. In addition, the expert from Marie Stopes, a private women’s health clinic providing abortion services, corroborated this point by saying that girls as young as nine have come in for abortions.

Women shared that while some men were against abortion, others forced their girlfriends to get abortions. Six participants stated that the biological fathers or 'baby daddies' often pressure their girlfriends to terminate their pregnancies. Interviewees shared that men without the money to raise a baby or in multiple secret relationships would usually pay or pressure women to get abortions. A 28-year-old participant stated, "I know one of my friends, he was giving her money to go and do an abortion. She couldn't do it. Instead, she 'went' and did a scan." Despite the woman choosing not to go through with the abortion, she was under immense pressure to terminate the pregnancy against her wishes. Similarly, Participant 20 shared a story about a 23-year-old woman who aborted a baby at 7-8 months. The community found out and made statements like, "She did what? Abortion? How could she kill an innocent child. She should be jailed." However, the participant gathered further information, "At that time, I was friends with her sister. She told me no, the boyfriend did not want to have the child. The boyfriend was putting pressure on her. He will do something, or he will kill her and the child. So she didn't have a choice. But most people didn't know that...No at first she didn't want to do it. The baby daddy was putting too much pressure. And the boyfriend was married." The community's judgment was directed towards the woman; however, they did not know the whole story. She would never have gotten an abortion had her boyfriend not forced her into the decision.

Finally, fear of familial judgment was the final driving factor of abortions described by the women in the Cato Manor sample. Eight participants stated that fear of parents drove women to get abortions. Participant 23 stated that women mostly get abortions because they're scared of their parents and families. Another participant further explained, "Obviously, you break your virginity, fine, you can still hide that. But then, when you get pregnant. What will my parents think? My parents will disown me. And with us, if for say my parents disown me. It's not like maybe with you guys where as your mom is maybe not going to talk to you. The whole family." Other participants also stated they could be disowned or kicked out of their childhood homes. The fear of family was further exemplified by a translated story shared by Participant 31, a 65-year-old woman, regarding a woman who died from an illegal abortion due to fear of the community's reaction in the late 1990s. She shared,

"There's a lady that she knew from the Mblungo family, and she fell pregnant. During that time, she was scared of her parents. Because back then, it was such a shame. It was not popular. You did it secretly. It wasn't legal at that time. She went and had this abortion. Another lady did it, and she didn't have the proper equipment. And her uterus got an infection. And it was eating her and eating her. It was spreading and spreading until she died...she never told the baby daddy. And the baby daddy was like, my mom is a nurse, and we were never not going to take responsibility for the child. If only she had told me, she wouldn't be dead."

Women have died even after abortion was legalized due to fear of parental reactions. While countless additional factors contribute to a woman's choice to abort, these three factors were most frequently mentioned within the sample population.

"The families and the baby daddies"

The factors inducing women to get abortions directly conflict with the negative cultural and religious shame women who openly access abortion services face. Women are culturally and religiously expected to disapprove of abortion, yet many need to have an abortion for various reasons. Therefore, they struggle deciding whether to follow their cultural and religious moral compass or abort a baby they do not have the means or support to raise. However, in cases of teenage pregnancy, pressure from baby daddies, and familial fear, the need to get an abortion often outweighs the socio-cultural drawbacks ingrained in the community. As seen earlier, the older generation is judgmental of the younger for having "careless" and unprotected sex resulting in abortions, while the younger generation is more forgiving. Regardless of judgment, both generations agree that teenage pregnancy is a problem in the community. As cited in one study, the two main factors leading to abortion are 'wanting to focus on studies' and 'not being ready to be a parent.'¹⁷ The desire to continue education and the sense of unpreparedness for parenthood can be associated with teenage or early pregnancy. Being young and unable to provide for oneself, let alone a family is a significant inducement to terminate a pregnancy. Many teenagers and young adults are still in school, and the responsibility of concurrently raising a child is too much. Despite their community's cultural and religious beliefs, the need to finish an education outweighs the potential shame of abortion.

A similar study was conducted in urban Mozambique by Frederico, Mónica, *et al.*, a country with less liberal but similar abortion laws to South Africa. However, the factors they identified as driving abortion differed significantly from this study. They cited that lack of autonomy, knowledge, poor availability of facilities, and provider influence led to abortions.¹⁹ Only lack of autonomy was relevant among this study's sample population, as participants mentioned no others. Lack of autonomy was described as pressure from family or baby daddies, two of the main factors described by these participants. The context of South Africa and the unique socio-cultural attitudes in Cato Manor must differ from Maputo, Mozambique. Abortion was legalized in Mozambique before 12 weeks in 2014.²⁰ Women in Cato Manor have access to abortion education and facilities due to the more liberal South African Constitution and earlier legalization date. Despite being just across the border, Mozambican women face different challenges than South African women. However, hierarchies of power and the struggle of a patriarchal society exist within all modern societies. Therefore, it is unsurprising that lack of autonomy was an overlapping factor within both studies.

Participants described 'baby daddies' as forcing their partners to get abortions. Abortion can be a result of unequal power dynamics in relationships. Men are described as giving money or threatening their partners to get an abortion. A surprising story was shared by Participant 20 in the findings above. The impact of partner pressure resulted in a late-term abortion against the woman's will. She did not want to abort her baby, yet she worried for her life. Fear as a motivator to get an abortion outweighed any potential worry of community shame or commitment to personal attitudes towards abortion. Literature within low and middle-income countries, such as Strong, supports participant testimonies of experiencing pressure from the biological father. Strong states that men use money and resources to control women's access to abortions. This data shows that women are shamed for abortions due to cultural and religious beliefs seen in earlier sections, while men are often equally, if not more, to blame.

Pregnancy cannot be hidden from one's parents and shows that the woman has broken traditional norms of delayed sexual debut. Therefore, rather than admit to premarital sex, some women will get an abortion out of fear of their parents' response. Fear connects many factors resulting in abortion, as it dominates cultural and religious expectations. Familial pressure is not often cited in literature; however, in the dense and close-knit community in Cato Manor, children may live with parents into adulthood. As a result, pregnancy is difficult to hide, and abortion may be viewed as a solution.

CONCLUSIONS

The research revealed the complex and dissonant factors pushing and pulling women from terminating a pregnancy. Zulu culture persists in Cato Manor, with strict spiritual views about abortion. Women fear bad luck and disastrous repercussions from their ancestors after an abortion. Many revealed that Zulu rituals are still practiced and passed down among community members if abortion ever occurs. As a result, abortion is not accepted in Zulu culture and remains a significant factor in women's choice of whether to receive an abortion. Religious judgment further contributes to the negative attitude towards abortion in Cato Manor by overlapping and augmenting cultural judgment. Most religious participants labeled abortion as 'murder' and 'a sin' and found it so abominable they disliked discussing it. Participants agreed that the combination of religion and culture were the two most significant factors resulting in abortion shame and negative attitudes. However, intergenerational change due to modernization and education has changed perceptions of pregnancy termination in the community. Younger participants were more educated about abortion and open to understanding the struggles of their peers seeking abortions. In contrast to the religious and cultural barriers to abortion, young and unplanned pregnancy, partner compulsion, and familial pressure have left women with no choice but pregnancy termination. Clear in the Cato Manor community were the conflicting issues of shame and reasons abortions were needed. Women were informed by cultural and religious expectations that abortion was morally reprehensible. However, life within their society and community often left them with no choice but to terminate their pregnancy. This internal conflict forces women to turn to unsafe abortion, despite easily accessible legal options. Rather than face community judgment, women opt for the private backstreet option. Socio-cultural pressures have a tremendous effect on the health of women who are dying due to unnecessary, illegal abortions. Related to this problem is socio-economic and gender-based discrimination resulting from a patriarchal and post-colonial society. To prevent women from needlessly dying from backstreet abortions the hierarchical social fabric of South Africa must be interrupted. Abortion is legal in South Africa, yet it will not be utilized by the population most in need unless socio-cultural expectations change to accommodate legal pregnancy termination.

Recommendations for further study include digging into the relationship between culture and religion and its impact on abortion. These participants were never directly asked how Zulu culture and Christianity overlap to create negative perceptions of abortion. More research should be done in this area to understand where judgment arises. In addition, research should be done among a generation younger than 20-30 regarding abortion attitudes. Much has changed in the last 20 years, and it would be helpful to understand teenagers' perspectives on abortion and if they continue the trend of acceptance seen among the younger generation. Lastly, this research reveals why women get backstreet abortions, yet more could be done on the first-hand experience of getting an illegal abortion. The context of this community research would be helpful to build off when analyzing personal stories of unlawful abortions. It would be interesting to determine whether the community is correct in its assumptions for the reasons women get illegal abortions. Finally, and most critically, research is necessary on how to lower rates of unsafe abortion in South Africa. The critical problem of backstreet abortions and why women get them has been identified in this study, yet no solution currently exists to prevent further maternal mortality. Legalization was the first step, yet research must be done to get women to utilize their rights to safe abortion.

ACKNOWLEDGEMENTS

How lucky I am to have a wonderful family who supported my choice to travel to the other side of the world and continued to encourage me throughout every frantic phone call along the way. To my mother Diana who inspires me to entrust the goodness she shares with me with each person I've met, and to persevere through adversity while still pausing to enjoy the small moments along the way. To my father Andrew who has forever been the North Star to my moral compass and my reliable source of advice. And to Audrey, my little sister and best friend, for being my constant support system and source of ceaseless laughter. Thank you

all for supporting my adventures, and for creating a home which will always be the safe space I return to. Little did I know how much you prepared me to listen and learn in a new community, oceans away from home.

I also must thank my South African family who welcomed a strange looking American to share their dinner table. To my Mama with a cold glass of juice waiting after each tiring day of classes. And to Aphiwe and Siyamuthanda for the joy and laughter their dancing skills brought each evening. And lastly a thank you to NThandokazi for being my second sister and much-needed isiZulu tutor. You opened your house, and it became my home. I'll never be able to thank you enough for the love you showed a stranger.

Finally, to our wonderful instructors John McGladdery and Christine McGladdery. Thank you, John, for challenging my world perspectives at every turn and assisting me throughout the writing and editing process. Christine, your advice on research methods and maintaining the highest level of ethical consideration was invaluable. Finally, NoThando none of this would have been possible without the access and insight you provided into Cato Manor. Thank you for setting up endless interviews and for your help each step of the way.

REFERENCES

1. Bankole, *Fact Sheet Abortion in Africa - Guttmacher Institute*, https://guttmacher.org/sites/default/files/factsheet/ib_awnw-africa.pdf (accessed Apr 2023)
2. The World's Abortion Laws, Center for Reproductive Rights, reproductiverights.org/maps/worlds-abortion-laws/ (accessed Apr 2023)
3. Hodes, R. (2016) The culture of illegal abortion in South Africa, *Journal of Southern African Studies* 42.1, 79-93. <https://doi.org/10.1080/03057070.2016.1133086>
4. Goldberg, J. (January 26, 2021) UN Human Rights Committee Asserts That Access to Abortion and Prevention of Maternal Mortality Are Human Rights, Center for Reproductive Rights, <https://reproductiverights.org/un-human-rights-committee-asserts-that-access-to-abortion-and-prevention-of-maternal-mortality-are-human-rights/> (accessed Apr 2023)
5. WAZIMAP Profile: Ward 29 (59500029), Ethekwini, KwaZulu-Natal. *Wazimap*, <https://wazimap.co.za/profiles/ward-59500029-ethekwini-ward-29-59500029/#demographics> (accessed Apr 2023).
6. McGill, J. (2006) Abortion in South Africa: how we got here, the consequences, and what is needed, *Journal for Christian Science*, Special ed. 2, 195-222.
7. Mhlanga, R. E. (2003) Abortion: developments and impact in South Africa, *British medical bulletin* 67.1, 115-126. <https://doi.org/10.1093/bmb/ldg006>.
8. Choice on Termination of Pregnancy Act 92 of 1996, Republic of South Africa - Gov, https://www.gov.za/sites/default/files/gcis_document/201409/act92of1996.pdf (accessed Apr 2023).
9. Johnston, R. Historical Abortion Statistics, South Africa, www.johnstonsarchive.net/policy/abortion/ab-southafrica.html (accessed Apr 2023)
10. du Plessis, Sofika, Macleod, and Mthethwa (2019) Assessing the Impact of the Expanded Global Gag Rule in South Africa, *Critical Studies in Sexualities and Reproduction Rhodes University*, <https://srjc.org.za/2019/03/25/assessing-the-impact-of-the-expanded-global-gag-rule-in-south-africa/> (accessed Apr 2023)
11. Jooe, F. (2018) Is there room for religious ethics in South African abortion law? *South African Journal of Bioethics and Law* 11.1, 46-51. <https://doi.org/10.7196/SAJBL.2018.v11i1.00624>
12. Jarvis, J., Mthiyane, N. P. (2019) Exploring Religious and Cultural Identities and the Right to Bodily Self-Determination in a South African Higher Education Context, *African Journal of Gender and Religion*, 25(1). <https://doi.org/10.14426/ajgr.v25i1.889>
13. Mosley, Anderson, Harris, Fleming, and Schulz (2020) Attitudes toward abortion, social welfare programs, and gender roles in the US and South Africa, *Critical public health* 30.4, 441-456. <https://doi.org/10.1080/09581596.2019.1601683>
14. Albertyn, C. (2015) Claiming and defending abortion rights in South Africa, *Revista Direito GV* 11, 429-454. <https://doi.org/10.1590/1808-2432201519>
15. Mosley, King, Schulz, Harris, De Wet, and Anderson (2017) Abortion attitudes among South Africans: findings from the 2013 social attitudes survey, *Culture, health & sexuality* 19.8, 918-933. <https://doi.org/10.1080/13691058.2016.1272715>
16. Nilsson, B., Edin, K., Kinsman, J., Kahn, K., and Norris, S. (2020) Obstacles to intergenerational communication in caregivers' narratives regarding young people's sexual and reproductive health and lifestyle in rural South Africa, *BMC public health* 20, 1-11. <https://doi.org/10.1186/s12889-020-08780-9>
17. Ramiyad, D. and Patel, C. J. (2016) Exploring South African adolescents' knowledge of abortion legislation and attitudes to abortion: Sexual status and gender differences. *South African Journal of Child Health*, 10(2), 105-106. <https://dx.doi.org/10.7196/sajch.2016.v10i2.1137>
18. Masanabo, K. K. D., Govender, I., and Bongongo, T. (2020) Reasons women terminate their pregnancies legally and their contraceptive practices at Soshanguve 3 Community Health Centre, Tshwane district, South Africa, *South African Family Practice* 62.2, e1-e5. <https://doi.org/10.4102/sajfp.v62i1.4310>

19. Frederico, Michielsen, Arnaldo, and Decat (2018) Factors influencing abortion decision-making processes among young women, *International journal of environmental research and public health* 15.2, 329. <https://doi.org/10.3390/ijerph15020329>
20. Mozambique, Ipas, www.ipas.org/where-we-work/africa/mozambique/ (accessed Mar 2023)
21. Strong, J. (2022) Men's involvement in women's abortion-related care: a scoping review of evidence from low-and middle-income countries, *Sexual and Reproductive Health Matters* 30.1, 1-25. <https://doi.org/10.1080/26410397.2022.2040774>
22. Chemlal, S. and Russo, G. (2019) Why Do They Take the Risk? A Systematic Review of the Qualitative Literature on Informal Sector Abortions in Settings Where Abortion Is Legal, *BMC Women's Health* 19.5, 1-11. <https://doi.org/10.1186/s12905-019-0751-0>
23. Kaswa, R. (2021) Barriers to access second-trimester abortion: A case report, *South African Family Practice* 63.1, 1-3. <https://doi.org/10.4102/sajfp.v63i1.5028>
24. Meel, B. L. (2022) Why Do Women Not to Go for Abortion in a Designated Legal Abortion Facility in Transkei Region of South Africa?, *Indian Journal of Forensic Medicine & Toxicology* 16.1, 1536-1541. <https://doi.org/10.37506/ijfnt.v16i1.17720>
25. Netshinombelo, M., Maputle, M. S., and Ramathuba, D. U. (2022) Women's Perceived Barriers to Accessing Post-Abortion Care Services in Selected Districts in KwaZulu Natal Province, South Africa: A Qualitative Study, *Annals of Global Health* 88.1, 75. <https://doi.org/10.5334/aogh.3888>
26. Harries, Daskilewicz, Bessenaar, and Gerdtts (2021) Understanding abortion seeking care outside of formal health care settings in Cape Town, South Africa: a qualitative study, *Reproductive Health* 18, 1-8. <https://doi.org/10.1186/s12978-021-01243-3>
27. Percy, W. H., Kostere, K., and Kostere, S. (2015) Generic Qualitative Research in Psychology, *The Qualitative Report* 20.2, 76-85. <https://doi.org/10.46743/2160-3715/2015.2097>
28. Hennink, M. and Kaiser, B. N. (2022) Sample sizes for saturation in qualitative research: A systematic review of empirical tests, *Social Science & Medicine* 292, 114523. <https://doi.org/10.1016/j.socscimed.2021.114523>
29. Anderson, C. and Kirkpatrick, S. (2015) Narrative Interviewing, *International Journal of Clinical Pharmacy* 38.3, 1-4. <https://doi.org/10.1007/s11096-015-0222-0>

ABOUT STUDENT AUTHOR

Chloe Sachs will graduate in 2024 from Washington University in St. Louis with a bachelor's degree in History and Biology. After graduation she will go on to attend medical school to prepare for a career as a pediatric surgeon. Chloe has completed additional published research on Mental Health in Korean History and is currently working on a Senior Honors Thesis on the role of female medical practitioners in the Chosen dynasty. She primarily conducts research within the realm of history and healthcare among marginalized groups. This study was conducted for the Community Health and Social Policy program in South Africa through the School for International Training.

PRESS SUMMARY

Abortion is legal in South Africa, yet maternal mortality remains high due to unsafe abortion. In post-Apartheid South Africa, a history of oppression, differing cultural norms, and varied religious beliefs all impact access to legal pregnancy termination. This study looks at the reasons why women still seek out backstreet abortions when legal options are available in the township of Cato Manor in KwaZulu-Natal. Through one-on-one personal interviews, narratives and trends were explored to understand how community beliefs and societal norms impact abortion attitudes and choice. Beliefs towards abortion remain negative in Cato Manor, resulting in women seeking out illegal abortions to maintain privacy and avoid judgment.

Dynamic Structural Equation Models: Promising Yet Concerning

Suryadyuti Baral* & Patrick J. Curran

Department of Psychology and Neuroscience, University of North Carolina - Chapel Hill, Chapel Hill, NC

<https://doi.org/10.33697/ajur.2023.096>

Student: sonnet@email.unc.edu*

Mentor: curran@unc.edu

ABSTRACT

Dynamic Structural Equation Model (DSEM) is a powerful statistical modeling approach that has recently gained popularity among researchers studying intensive longitudinal data. Despite its exciting potential, the stability and replicability of DSEM is yet to be closely examined. This study empirically investigates DSEM using recently published data to explore its strengths and potential limitations. The results show that while some of its parameter estimates are stable, others are characterized by substantial variation as a function of seemingly innocuous initial model estimation conditions. Indeed, some parameters fluctuate between significance and non-significance for the same model estimated using the same data. The instability of DSEM estimates poses a serious threat to the internal and external validity of conclusions drawn from its analyses, challenging the reproducibility of findings from applied research. Given the recent focus on the replication crisis in psychology, it is critical to address these issues as the popularity of DSEM in psychological research continues to rise. Several potential solutions are investigated to address this problem and recommendations of best practice are offered to applied researchers who plan to use DSEM in intensive longitudinal data analysis.

KEYWORDS

Dynamic Structural Equation Model; Bayesian; Robust Estimation; Intensive Longitudinal Data

INTRODUCTION

Human psychology weaves together like an intricate tapestry, where the threads of cognition and behavior intertwine, creating a rich and complex individual. One of the most important goals in the behavioral sciences is to disentangle the threads and study the cause—effect pathways that shape the individual. In order to study such relationships, researchers need longitudinal data—that is, repeated assessments of individuals collected over the course of weeks, months or even years. With increasing recognition that life unfolds continuously over time, there has been a push towards intensive longitudinal data (ILD) analysis in which a large number of assessments are taken in shorter time intervals of days or even hours.¹ Although characterized by many strengths, ILD comes with its own set of unique challenges in measurement and modeling. A key challenge is to satisfactorily examine co-developmental processes where cause-and-effect influences are studied in two or more processes over multiple time points. Quantitative models like autoregressive cross-lagged (ARCL) have been developed for the analysis of longitudinal data, although it is increasingly appreciated that such traditional models fail to accurately capture co-developmental processes in ILD.² The tapestry of human psychology, quite unsurprisingly, is just too tightly knit and it has challenged the statistical might of generations of quantitative psychologists.

However, a novel method of modeling these dynamic processes has recently been introduced for the analysis of ILD: Dynamic Structural Equation Model, or DSEM.³ DSEM has the potential to move well beyond the confines of conventional techniques. It provides a comprehensive framework for modeling and analyzing the reciprocal interplay of co-developing phenomena at both the individual and group level. Indeed, DSEM offers the prospect of testing research hypotheses in the behavioral sciences in ways not previously possible. DSEM's application in the myriad fields of social sciences has garnered significant attention. A recent study from McNeish and Hamaker⁴ is an illustrative discussion on how DSEM can be employed in applied research. Yet, given DSEM's recent development and strong encouragement for use, almost nothing is known about the stability of its estimation and the replicability of its results. It is paramount that DSEM be subjected to rigorous scrutiny before its widespread adoption in applied research. The goal of our paper is to provide an initial examination of the stability and replicability of DSEM under conditions commonly encountered in applied research. We reanalyze previously published data across a range of initial conditions in order to demonstrate what aspects of the DSEM are and are not stable and replicable.

METHODS AND PROCEDURES

DSEM brings together elements of three well-established analytic methods: Structural Equation Modeling (SEM), multilevel modeling (MLM) and time series analysis.^{3,4} Similar to MLM, DSEM accommodates nested structures in the data, where observations are clustered under higher level units (here, time nested within individuals). The model representing the higher-level unit is called the *between-person* model and the individual models nested under the higher-level unit is called the *within-person* model. DSEM uses time series analysis to model the autocorrelations in the within-person model for intensive longitudinal data. Moreover, incorporation of SEM allows DSEM to model the individual differences in the time-series parameters as latent variables. This integrative approach fills critical gaps left by individual methods. For instance, MLM handles interindividual variability but not latent variables. SEM deals with latent variables but lacks the flexibility for highly dense time measures and for random effects on model parameters. By bringing these methodologies together, DSEM emerges as a powerful tool capable of addressing the limitations of its constituent methods and offering a more comprehensive analytical framework for studying complex data structures.

DSEM provides both a powerful and flexible statistical framework, making it a valuable tool in the social sciences. A clinical psychologist interested in nicotine addiction and depression may hypothesize that depression and urge to smoke are entangled together in a codeveloping process.^{5,6} Job stress and home stress may act as a catalyst to the situation by elevating an individual's level of depression and in turn entangling the urge to smoke and depression in a tighter yarn. McNeish and Hamaker⁴ provide a guide to any researcher who wishes to test out such hypotheses using DSEM. The simulated dataset from the paper had measurements on depression, urge to smoke, home stress and job stress for 100 individuals across 50 discrete time points. This dataset is suitable to be analyzed using DSEM because it has a comparatively large sample of individuals, all measured over multiple time points. The presence of an observed variable with an autoregressive component (in case of urge to smoke) and without the autoregressive component (in case of depression) requires time series analysis. Proper treatment of the time invariant covariates of job and home stress necessitates the use of MLM. The latent variables used in the model draw heavily from the SEM literature. In short, the complexities of the dataset demand a flexible framework that attends to all these needs. Here we reanalyze the same data with our primary focus on the time-invariant covariate (TIC) DSEM as elucidated in McNeish & Hamaker.⁴

The path diagrams and representative equations¹⁻⁶ delineate the TIC DSEM. In this context, depression and urge to smoke are defined as time-varying covariates (or TVCs). For each individual, the urge to smoke at a time point (t) is dependent on their urge to smoke at the previous time point (t-1) and their depression at the same time point (t). This is represented in **Equation 1**. The random slopes at the *within-person* level resurface in the consequent **Equations 2-4**. At the *between-person* level, job stress and home stress affect the random slopes. Job and home stress are constant for each individual and therefore act as the TICs. The TICs are grand mean centered as opposed to Dep,^b which is latent person-mean centered. In **Figure 1a**, the gamma (γ) variables represent the fixed effect of its predictor, and the u terms represent the random inter-subject effect. The tau (τ) variables stand for variances of the u parameters. For the physical description of each parameter, refer to **Table 1**.

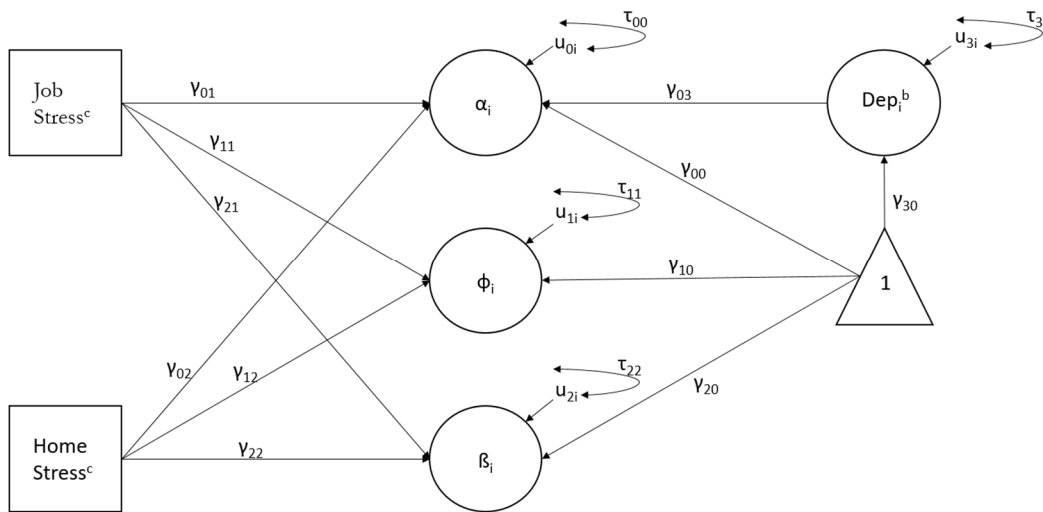


Figure 1a.

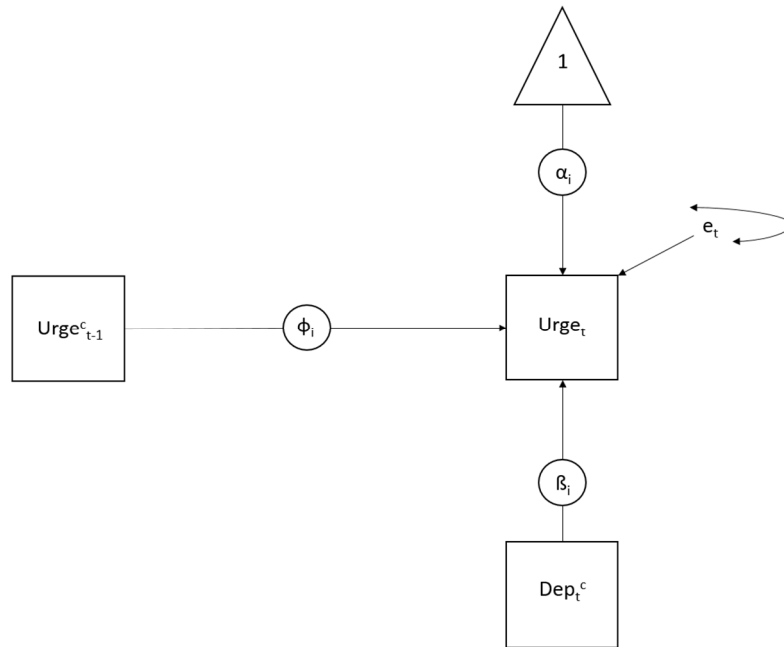


Figure 1b.

Figure 1. The figure displays the path diagrams of the between-person model (1a) and the within-person model (1b) for the TIC DSEM outlined in McNeish and Hamaker.⁴ The rectangular boxes stand for variables that were directly measured and the circles represent latent variables estimated from the data. Note: path diagrams have been adapted from McNeish and Hamaker.⁴

$$Urge_{ti} = \alpha_i + \phi_i Urge_{(t-1)i}^c + \beta_i Dep_{ii}^c + e_{ti} \tag{Equation 1.}$$

$$\alpha_i = \gamma_{00} + \gamma_{01} JobStress_i^c + \gamma_{02} HomeStress_i^c + \gamma_{03} Dep_i^b + u_{0i} \tag{Equation 2.}$$

$$\phi_i = \gamma_{10} + \gamma_{11} JobStress_i^c + \gamma_{12} HomeStress_i^c + u_{1i} \tag{Equation 3.}$$

$$\beta_i = \gamma_{20} + \gamma_{21} JobStress_i^c + \gamma_{22} HomeStress_i^c + u_{2i} \tag{Equation 4.}$$

$$Dep_{ii}^n = Dep_{ii}^c + Dep_i^b \tag{Equation 5.}$$

$$Dep_i^b = \gamma_{30} + u_{3i} \tag{Equation 6.}$$

Including the latent person-mean centered Dep_{ii}^c in the within-person model and the latent person-mean Dep_i^b in the between-person model allows the complete disaggregation of the total effect of depression observed in the raw data (Equation 5). Latent person-mean centering approach treats Dep_i^b as an unknown quantity that has to be estimated and thus properly accounts for measurement error.⁷ This inclusion allows researchers to discern the impact of unit change in a covariate on its outcome at a specific measurement occasion (the within-person effect). In this scenario, Dep_i^b influences α_i (Equation 2), the mean urge to smoke of an individual which also shows up at the within person level. Moreover, researchers can inspect how a one-unit change in the covariate mean across all measurement occasions affects the average of the outcome variable (the between-person effect). By including both effects in the model simultaneously, it is possible to investigate whether the within-person and between-person effects differ.

While DSEM is a versatile framework, it is quite complex and requires advanced methods for statistical estimation. The commercial software program Mplus,⁸ the only package that currently implements DSEM, uses Bayesian estimation with a Gibbs sampler. The Gibbs sampling approach enables the estimation of parameters using conditional distributions, given that the conditional distributions are known and are easier to sample from compared to the unknown and often complex joint distribution.⁹ It begins with an initial seed value for all the parameters. In the first iteration, it fixes the values of all parameters besides one and samples the value of the unfixed parameter from its conditional distribution. The estimator then selects a separate parameter, fixes the rest and samples from its conditional distribution. This process goes on until a new sample of values for all the parameters have been generated from their respective conditional distributions. The estimator continues sampling in multiple

iterations and the samples begin to approximate the joint distribution of the parameters. This technique is elegant because it uses the local dependencies and enables efficient exploration of high-dimensional spaces without having to directly sample from the often intricately complex joint distribution. In a model like DSEM with multiple parameters and a complex parameter space, Gibbs sampling is a logical option for estimation.

However, Gibbs sampling is not without its own set of challenges.^{10,11} One key concern lies in its sensitivity to initial conditions, or the *seed values*. Since initial conditions guide its sampling of the posterior distribution, the first few samples can be biased, sometimes substantially so. Therefore, the initial samples are often discarded, a process referred to as the *burn-in*. In addition to burn-in, multiple sampling chains can be used to explore the space. The initial conditions assigned to the parameters might push the chains towards a local region of the distribution, making samples from the distribution seem biased. This underscores the importance of an extensive number of iterations to facilitate convergence.

To assess convergence, a commonly employed criterion is the Potential Scale Reduction Factor (PSRF)¹². This criterion hinges on the idea of interchain variability, wherein a PSRF value equal to 1 means that the samples acquired from the chains are indistinguishable from each other. From a PSRF=1, it can be assumed that the chains are sampling from the same distribution—the true posterior distribution. McNeish and Hamaker⁴ adhered to the default values within Mplus,⁸ in which the first half of the samples were discarded as burn-in and two chains were used in exploration. The number of iterations was set to a maximum of 1000 and the convergence criteria called the Potential Scale Reduction Factor (PSRF) was set to the default value of 1.1.

While a PSRF value of 1 may take an unfeasibly long time to achieve, PSRF values close to 1 are achievable. However, this convergence criterion does not ensure that the samples obtained from the chains generate reproducible samples of the posterior distribution. It is possible that the stochastic algorithm becomes stranded in some region of the parameter space and requires a substantially longer period of time to converge to the true posterior distribution.¹³ Moreover, PSRF is extremely dependent on the shape of the posterior. It has been observed that for heavily skewed posterior distributions PSRF does not converge to 1.0 even with increasing sample size.¹⁴ Therefore, it is essential to use multiple random seeds or initial conditions to check if the chains are producing similar or the same parameter estimates. If, after a reasonably extensive list of initial conditions, each independent of the other, the parameter values rest stably at the same values, it can be said that the estimator is sampling from the true posterior distribution. Such assessment of parameter stability is critical while investigating model stability.

McNeish and Hamaker only used a single seed value and a single PSRF value for all of their analyses. It is unknown if their final solution was dependent on the initial conditions of chains and the default parameters of the convergence criteria. To assess the stability of the parameter estimates and their dependence on initial conditions, we first randomly sampled 1000 seed values (without replacement) from integers ranging from 1 to 100000 using R (version 4.2.2).¹⁵ Next, we estimated 1000 separate DSEMs, each using one of the randomly sampled seeds. The models were run on Mplus using MplusAutomation¹⁶ (version 1.1.0) in R. To further probe the stability of these estimates, we introduced variations in the estimator such that the maximum number of iterations allowed was systematically altered (values of 1000, 10000 and 30000) as well as the value of the PSRF (values of 1.1, 1.05, 1.01 and 1.005). These manipulations of the convergence criteria were aimed to provide a more nuanced understanding of the impact of such criteria and initial conditions in the estimation of DSEM. The parameter estimates, defined by default as the median of the parameter's posterior distribution, were recorded across the 1000 initial conditions encompassing the variations in the PSRF thresholds and iteration counts. The median values of each of the parameters for the 1000 different seed values were visualized through boxplots for each of the conditions. The analysis was done in R¹⁵ and the visualizations were created using ggplot2 (version 3.4.4).¹⁷ To reduce redundancy, for some of the conditions we report the mean, median, standard deviations and quartiles of the parameters instead of plotting the boxplots.

Besides the parameter estimates, a record was maintained concerning the inclusion of 0 in each parameter's credible interval in its posterior distribution. In Bayesian statistics, the posterior distribution of the parameters is used to discuss its estimate and precision. However, frequentists using Gibbs sampling to estimate DSEM might use the presence or absence of 0 in the credible intervals of the parameter distribution for significance testing. It should be noted that this is not a true null hypothesis significance test, as the concept of null hypothesis testing does not fit well in the pure Bayesian perspective. However, as seen in McNeish and Hamaker⁴, researchers use Bayesian credible intervals for null hypothesis testing. We do not wish to deviate from the analysis perspective used in McNeish and Hamaker⁴, therefore we use the inclusion of 0 in the Bayesian credible intervals as a test of significance. The following discussion will include both the stability of the parameter estimates and the significance of the parameters to foster a more holistic insight into the problem at hand.

RESULTS

We present the findings from the TIC DSEM analysis, as discussed in McNeish and Hamaker⁴ in Table 1.

Effect	Notation	Posterior Median	95% Credible Interval
Intercept (Alpha): overall intercept capturing the baseline urge to smoke	γ_{00}	.06	[-.14,.30]
Intercept (Phi): overall autoregressive intercept capturing the baseline autoregressive effect	γ_{10}	.19	[.16,.22]
Intercept (Beta): overall effect of depression on the urge to smoke capturing the baseline effect	γ_{20}	.79	[.62,.95]
Intercept (Dep): overall effect capturing the baseline influence of depression on the urge to smoke	γ_{30}	.02	[-.01,.05]
Alpha on Job Stress: effect of job stress on the baseline urge to smoke	γ_{01}	.50	[.35,.65]
Alpha on Home Stress: effect of home stress on the baseline urge to smoke	γ_{02}	.33	[.19,.46]
Alpha on Dep: effect of the person mean of depression on the baseline urge to smoke	γ_{03}	-2.35	[-11.22,4.35]
Phi on Job Stress: effect of job stress on the autoregressive effect	γ_{11}	.12	[.09,.15]
Phi on Home Stress: effect of home stress on the autoregressive effect	γ_{12}	.06	[.03,.08]
Beta on Job Stress: effect of job stress on slope of depression on urge	γ_{21}	.29	[.11,.48]
Beta on Home Stress: effect of home stress on slope of depression on urge	γ_{22}	.35	[.17,.51]
Var. (Alpha): individual specific baseline urge to smoke	τ_{00}	.34	[.16,.49]
Var. (Phi): variance of the autoregression	τ_{11}	.01	[.00,.01]
Var. (Beta): variance of influence of depression on urge	τ_{22}	.64	[.47,.88]
Var. (Dep): variance of depression	τ_{33}	.01	[.00,.01]
Res. Var. (Urge): residual variance of urge to smoke	σ^2	1.14	[1.09,1.18]

Table 1. Estimates and 95% Credible Intervals for the TIC DSEM. Note: The data have been taken from McNeish and Hamaker⁴ and these are estimates from one seed value with 1000 being the maximum number of iterations allowed for the estimator.

The analysis reveals significant time-invariant effects in the model. Covariates for the intercept predict that stresses in the work and home environment increase the baseline Urge to Smoke for individuals. Specifically, a one-unit increase in Job Stress leads to a .50 unit increase in the mean Urge to Smoke while one-unit increase in Home Stress leads to a .33 unit increase in the same. The estimates for γ_{11} and γ_{12} are non-null and indicate that Home Stress and Job Stress strengthen the carryover effects of Urge to Smoke by .06 and .12, respectively. γ_{21} and γ_{22} are significant as well. These parameters provide evidence that increases in Job Stress and Home Stress is similarly predictive of a stronger effect of Depression on Urge to Smoke. γ_{21} is estimated to be .29, which means a one—unit increase in Job Stress increases β_3 , or the slope of depression on urge, by .29. Similarly, γ_{22} estimated to .35 means a one unit change in Home Stress increases the slope of depression on urge by .35 units. Notably, the person mean of Depression does not appear to have any significant effect on the person mean of Urge to Smoke as γ_{03} has 0 in its 95% credible interval. This lack of a significant effect can be interesting to clinical psychologists because it implies that treating only depression will not have any effect on smoking tendencies and vice versa. This necessitates targeted interventions for both depression and smoking behavior. However, before clinicians begin developing novel interventions, it is imperative to check the robustness of these estimates to save time, effort and financial resources. Thus, we tested if McNeish and Hamaker’s reported estimates were a local solution using different seed values for the estimation.

Since the original paper had employed 1000 iterations as the convergence criterion, we began with the same. However, instead of using just one seed value, we initiated the sampler with 1000 *different* seeds. To our concern, a substantial subset of the seeds failed to converge. The fact that no valid results could be obtained from these instances was, in itself, a disconcerting finding. Therefore, we shifted our focus to the first 100 seeds that exhibited convergence. Figure 2 shows the estimates of the model parameters for the 100 initializations. Notably, the boxplots for most of the parameters show minimal variability indicating no sensitivity to the initial conditions of the sampler, affirming DSEM’s reliability in most instances. However, a concerning degree of variability was observed in γ_{03} , the effect of the person mean of depression on the baseline urge to smoke, exhibiting marked sensitivity to the initial conditions of the sampler. While γ_{30} , or the mean value of depression, is not as variable as γ_{03} , three seeds among the 100 emerged as statistically significant. These findings indicate that key substantive conclusions from the TIC DSEM could be fully

dictated by the arbitrary seed value chosen for the analysis. The challenges of non-convergence and the parameter instability underscore the need for a higher number of iterations in the Gibbs sampler to achieve more stable estimates.

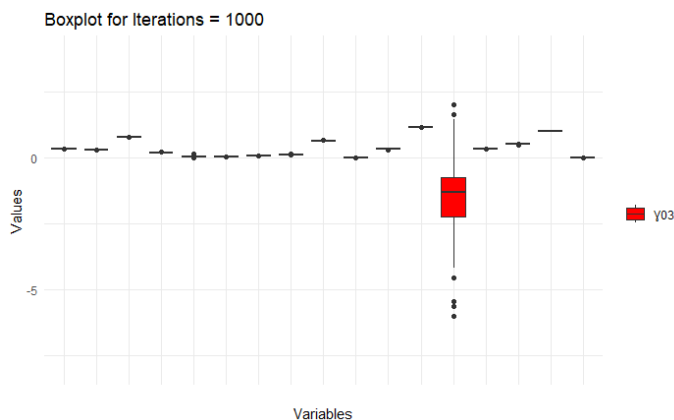


Figure 2. The figure displays boxplots for all TIC DSEM parameters based on 100 seeds, each subject to a 1000-iteration convergence criterion. γ_{03} clearly has the highest variability of all the parameters.

We extended the number of iterations substantially beyond the original recommendation and employed a convergence criterion of 10000 iterations. Under this criterion, all the seeds converged but the stability of parameter γ_{03} remained elusive. As seen in Figure 3, γ_{03} in this condition shows more variability than any of the other parameters. Moreover, 47 out of 1000 seeds yielded significant results for the mean value of depression (γ_{30}). This prompted an exploration of the convergence criterion of 30000 maximum iterations. The results persisted. γ_{03} showed no change in standard deviation, as can be seen in **Table 3**. γ_{30} had an equal number of significant seeds. These results concern us about potentially misleading significant results. If a random seed value can produce significance, researchers need to be cautious while interpreting results from DSEM.

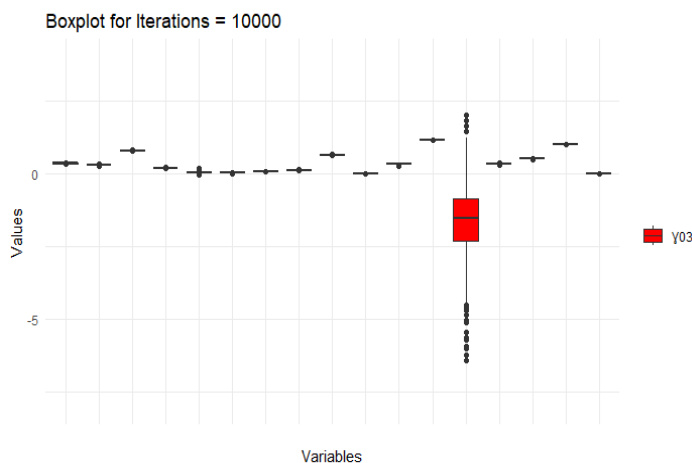


Figure 3. The figure displays boxplots for all TIC DSEM parameters based on 1000 seeds, each subject to a 10000-iteration convergence criterion. γ_{03} clearly has the highest variability of all the parameters. Note: This figure differs from Figure 2 on the crucial fact that for iterations set to 10000, all the models converged successfully, unlike iterations set to 1000. Therefore, this figure represents estimates from 1000 models as opposed to 100 from Figure 2.

Upon a thorough investigation into the convergence threshold used in the sampling algorithm, it became evident that the estimator relies on the similarity of samples from the posterior distribution. Using a higher number of iterations might result in the chains exploring the distribution better, but it does not ensure that the samples from the different chains will emerge similar to each other. Therefore, we employed a convergence threshold based on interchain variability called the Potential Scale Reduction Factor (PSRF). As described earlier, this criterion ensures that the estimator continues running until a predetermined value is attained.

The PSRF criterion is based on Gelman and Rubin’s seminal paper of 1992, where they recommended PSRF values less than or equal to 1.1 as an indicator of convergence. Consequently, Asparouhov and Muthén¹⁸ suggested that PSRF values between 1.1

and 1.05 can render samples from chains virtually indistinguishable for most models. Most studies have followed these recommendations and use PSRF values between 1.1 and 1.05. Thus, we set the PSRF convergence criterion to values of 1.1 and 1.05. Under both convergence criteria, γ_{03} exhibited variability. As shown in Figure 4, PSRF \leq 1.1 produced results similar to the iterations criteria. The critical difference lay in the fact that with stricter PSRF values one could significantly reduce the variability in the unstable parameter. Notably, the variability was significantly reduced when using PSRF \leq 1.05 compared to PSRF \leq 1.1 (see **Table 2**). The number of significant seeds for γ_{30} was 82 under PSRF \leq 1.1, while it reduced to 47 under PSRF \leq 1.05. It may be inferred that for models as complex as DSEM, lower convergence thresholds like 1.05 are better than the traditionally recommended value of 1.1. However, the result produced by PSRF \leq 1.05 was identical to those yielded by the convergence criteria based on the number of maximum iterations. In order to improve the stability of the parameter estimates, lower convergence thresholds need to be used.

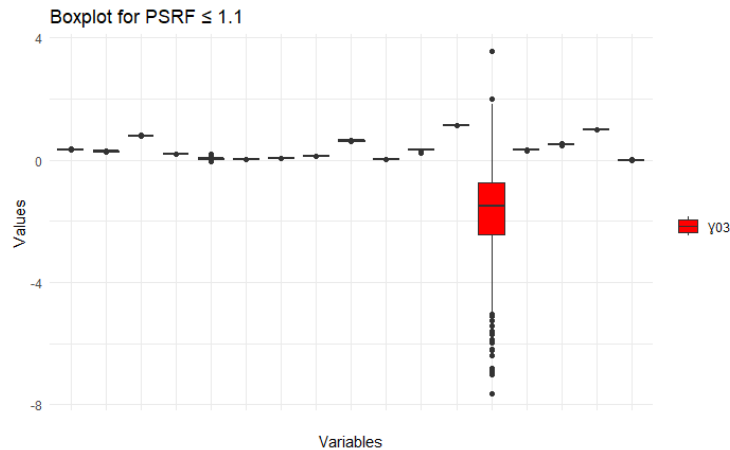


Figure 4. The figure displays boxplots for all TIC DSEM parameters based on 1000 seeds, each subject to a PSRF \leq 1.1 convergence criterion. γ_{03} clearly has the highest variability of all the parameters.

Lower convergence thresholds necessitate a departure from the standard values used with the PSRF criterion. Specifically, we tested thresholds as low as 1.01 and 1.005 to check if these bring stability to the parameter estimates. These values yielded a significant decrease in variability in the γ_{03} parameter as can be seen in **Table 2**. The number of significant seeds for γ_{30} dropped to 10 for PSRF \leq 1.01 and to only four seeds for PSRF \leq 1.005. The boxplots for the γ_{03} parameter under the seven different convergence criteria are shown in Figure 5 and it is apparent that the variability changes as a function of the convergence criterion. **Table 3** shows mean, median, standard deviation and the 25th and 75th percentiles of γ_{03} parameter for each condition of convergence. It reinforces the inference that the stricter PSRF criteria reduce the standard deviation of the unstable parameter. As noted earlier, PSRF \leq 1.05, iterations = 10000 and iterations = 30000 produce identical results. However, PSRF \leq 1.01 and PSRF \leq 1.005 produce significant reduction in variability. These results for the significance testing using pairwise Bartlett tests¹⁹ with Benjamini—Hochberg correction²⁰ are tabulated in **Table 2**. Although the PSRF \leq 1.005 produces estimates with significantly the least amount of variability, it should be noted that computational time was a greater challenge in the case for PSRF \leq 1.005 as each seed took about 50 seconds to converge with the 1000 seeds taking almost 14 hours. However, each seed took approximately 18 seconds to converge under PSRF \leq 1.01 which added up to 5 hours for the 1000 models. The substantial investment in time prompts a careful consideration on the part of the researcher while weighing the trade-offs of each convergence threshold. For this model, considering the modest differences between 1.01 and 1.005 thresholds while looking at the mean, median and variability values, it would be wise to stick with 1.01 as the best convergence threshold.

The above analysis sheds important light on some critical characteristics of the DSEM framework. While a majority of its parameter estimates demonstrate robustness to variations in estimator initializations, certain parameters are highly sensitive to seed values. This sensitivity impacts the estimates of certain parameters and the credible intervals of some others. For any researcher following the standard practice of using singular seed values to initialize parameters, such sensitivity in estimates would remain unknown and drastically influence the interpretations of results. To mitigate this sensitivity, we systematically altered the convergence criteria and were able to increase robustness in the parameter estimates. To ascertain statistical significance of the reduction in variability of the parameter γ_{03} , we employed Bartlett's test for equality of variances.¹⁹ To account for multiple comparisons, we used Benjamini—Hochberg false discovery rate correction.²⁰

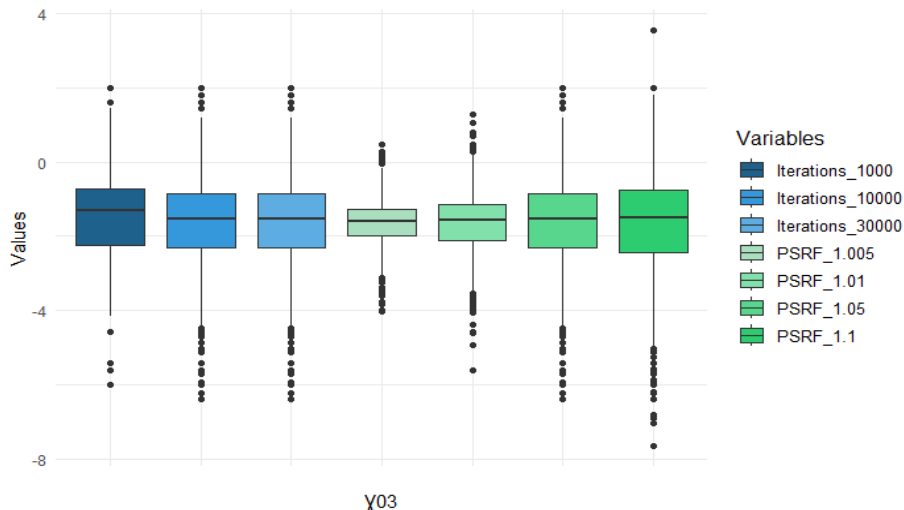


Figure 5. The figure displays boxplots for the γ_{03} parameter under all the 7 convergence criteria. The variability of the parameter changes as the convergence criteria changes with the least amount of variability being observed for $PSRF \leq 1.005$. Note: Iterations=1000 uses 100 models and their parameter estimates because many of the models failed to converge. All the other conditions allowed all the models to converge, therefore the other boxplots are made from estimates of 1000 models.

Group 1	Group 2	p value	p adjusted	sig
PSRF 1.1	PSRF 1.05	.00	.00	*
PSRF 1.1	PSRF 1.01	.00	.00	*
PSRF 1.1	PSRF 1.005	.00	.00	*
PSRF 1.1	Iterations 10000	.00	.00	*
PSRF 1.1	Iterations 30000	.00	.00	*
PSRF 1.05	PSRF 1.01	.00	.00	*
PSRF 1.05	PSRF 1.005	.00	.00	*
PSRF 1.05	Iterations 10000	1.00	1.00	
PSRF 1.05	Iterations 30000	1.00	1.00	
PSRF 1.01	PSRF 1.005	.00	.00	*
PSRF 1.01	Iterations 10000	.00	.00	*
PSRF 1.01	Iterations 30000	.00	.00	*
PSRF 1.005	Iterations 10000	.00	.00	*
PSRF 1.005	Iterations 30000	.00	.00	*
Iterations 10000	Iterations 30000	1.00	1.00	

Table 2. Pairwise Bartlett tests of the 6 groups with B—H FDR correction. Note: Iterations=1000 has not been included in this analysis due to the number of seeds that failed convergence.

PSRF 1.1	-1.66	-1.50	1.37	-2.45	-0.74
PSRF 1.05	-1.63	-1.53	1.15	-2.31	-0.86
PSRF 1.01	-1.65	-1.58	0.79	-2.10	-1.15
PSRF 1.005	-1.65	-1.61	0.60	-2.00	-1.26
Iterations 1000	-1.48	-1.32	1.36	-2.23	-0.73
Iterations 10000	-1.63	-1.53	1.15	-2.31	-0.86
Iterations 30000	-1.63	-1.53	1.15	-2.31	-0.86

Table 3. Mean, median, standard deviation, 25th and 75th percentiles for the estimate of γ_{03} parameter from 1000 models. For each condition, 1000 models with the same architecture but different initializations produced a posterior distribution for the γ_{03} parameter. The median of the posterior was treated as the estimate. Note: Iterations=1000 used 100 different seeds therefore the statistics come from the parameter estimates of those 100 models.

DISCUSSION

In this study, we explored the Dynamic Structural Equation Modeling (DSEM) framework with a focus towards the robustness of its parameter estimates. Using the TIC DSEM employed in McNeish and Hamaker,⁴ we discussed the various capabilities of the framework. As was shown in the aforementioned paper, DSEM has the ability to study time-invariant covariates in intensive longitudinal data. The model allowed an extensive analysis of how stresses in the work and home environment can have severe impacts on the carryover effects of the urge to smoke and in strengthening the relationship between depression and smoking tendencies. Using DSEM researchers can study such behavior better and thereby improve intervention programs.

However, our analyses unveiled a critical insight – not all of the model’s parameter estimates exhibited stability as a function of initial conditions. We used 1000 different seed values to initialize the Gibbs estimator. Across different initial conditions of the Gibbs sampler, a majority of the parameter estimates did not display variability. A few parameters proved to be exceedingly sensitive to seed values. The sensitivity affected the estimated values of certain parameters and the credible intervals of others. For the simulated dataset used in this study, the instabilities arose from γ_{03} , the effect of the person mean of depression on the baseline urge to smoke, and γ_{30} , or the mean value of depression. Unfortunately, it is difficult to make any inference as to why these parameters specifically show fluctuations. It is suspicious that both the variables are associated with the latent mean centered Dep, which may be the unstable component of the model. Here we can only ponder, but future research would do good to shed light on this issue. Although the underlying reasons are unclear, it should be acknowledged that this finding is disconcerting as model results are assumed to be stable across arbitrary initial conditions. Such lack of robustness can have cascading effects on the interpretation of results in academic papers. Yet all current recommendations for the widespread use of DSEM in practice do not address this issue. Therefore, it is imperative that applied researchers be aware of these issues before deciding to use DSEM in their research endeavors. Moreover, methods to reduce these instabilities need to be investigated.

To tackle the challenge of sensitivity to initial conditions of the Gibbs sampler, we systematically adjusted the convergence criterion. While McNeish and Hamaker⁴ used an upper bound on the number of iterations as the convergence criterion, our investigation revealed that transitioning to PSRF thresholds results in diminished parameter variability. Going beyond the traditional recommendations of PSRF thresholds between 1.1 and 1.05, we employed stricter convergence thresholds of 1.01 and 1.005. To assess statistical significance of the reduced variability in the parameter we used Bartlett’s test for the equality of variances. Benjamini—Hochberg false discovery rate correction was used to address multiple comparisons. $PSRF \leq 1.01$ and $PSRF \leq 1.005$ significantly reduced variability in the parameter estimate in comparison to $PSRF \leq 1.05$, as well as when the number of iterations was set to 10000 and 30000.

For a model as complex as TIC DSEM, which involves sampling from a complex posterior distribution, it is crucial to allow sufficient time for accurate exploration of the space. While more exploration of this complicated parameter space will allow more robust estimates, there is a trade-off between computational time and accuracy. Opting for a stringent threshold like 1.005 demands three times more computational time for convergence than a comparatively lenient threshold like 1.01. While the more stringent threshold will provide more robust estimates, the substantial increase in computational time makes it a less optimal choice. Consequently, we conclude that for the dataset at hand, a threshold value of 1.01 provides the best balance between robustness and computational efficiency. As DSEM gains traction in applied research, future investigators need to be mindful of these considerations and select the threshold that aligns best with their data.

CONCLUSION

This study explored Dynamic Structural Equation Models' robustness in parameter estimation, emphasizing its utility in studying time-invariant covariates within intensive longitudinal data. While DSEM provides a powerful tool for researchers to investigate complex behavioral dynamics, our paper reveals issues in the modeling framework that need to be accounted for. This was an initial investigation into the stability and replicability of DSEM on a simulated dataset. We discovered that a small subset of parameters is highly sensitive to initial conditions of the estimator. To address this lack in robustness and avoid local solutions, we propose transitioning to more stringent convergence thresholds. However, this comes at the cost of computational time. For the dataset used in this paper, a threshold of 1.01 is recommended to strike the best balance between robustness and efficiency. This underscores the necessity of researchers to be aware of the various challenges inherent to this novel framework.

While we have highlighted certain concerns within the DSEM framework, it needs to be emphasized that DSEM has immense potential in refining our understanding of dynamic human behavior as it unravels over time. Compared to other models in the literature like Autoregressive Cross Lagged (ARCL), DSEM does a better job in modeling co-developmental trajectories of psychological phenomena.^{3,21} We encourage researchers to use DSEM in their endeavors, but with a recommendation to adhere to best practices that help navigate the associated challenges. In light of our analyses, we propose prioritizing the PSRF convergence criterion over criteria involving the number of iterations of the Gibbs sampler. Additionally, smaller PSRF values (≤ 1.01) can yield more stable estimates, but the computational efficiency needs to be considered. Above all, we stress the importance of examining DSEM's estimates carefully before drawing conclusions. One may use different seed values for the estimation to test the robustness of the parameter estimates. However, there may be other and more effective ways of validating the results.

This study serves as an initial foray into the DSEM framework and is not meant to be comprehensive. There are a multitude of directions for future work some of which we mention here. We used a simulated dataset where the generating model was known. We acknowledge that our findings may be specific to this dataset, where the generating model was known. Future research would do well to explore simulations with different generating models to investigate if such issues consistently arise across datasets. Real world datasets, where the underlying model is unknown and complex, may present new challenges. Investigating if and how parameter instabilities show up under these real-world situations could yield valuable insights into the practical applications of DSEM. While this study unearthed the presence of instabilities in certain parameter estimates, we were not able to shed light on their underlying reasons. Simulations attempting to resolve these mysteries are highly recommended. Another promising direction of work lies in the exploration of priors for DSEM. We employed non-informative priors for our study. Future investigations could explore the sensitivity of parameters to mildly informative priors, as such adjustments might influence the posterior distribution, mitigate instabilities or improve computational efficiency. These are a few of the many avenues where more work needs to be done. Additionally, this study raised questions about the impact of Bayesian estimation in Mplus for the implementation of DSEM. Alternative software packages like JAGS/STAN should be considered to fit these models and test their robustness. At this point in time, DSEM can only be implemented in Mplus, opening up myriad opportunities for methodologists to build packages aimed at fitting these models. This is an ongoing quest for improvement that aims to empower researchers with a more robust and versatile tool to study human psychology. We hope that this study is just the first of many papers aimed at enhancing the DSEM framework.

ACKNOWLEDGEMENTS

The authors thank Dr. Daniel McNeish for providing access to the data from his paper which proved invaluable to this project. The guidance and support extended by the professors and graduate students at the L.L. Thurstone Psychometric Laboratory are sincerely appreciated. The authors express their gratitude towards UNC's Department of Psychology and Neuroscience for providing the financial support in acquiring the Mplus software license.

REFERENCES

1. Dunton, G. F., Rothman, A. J., Leventhal, A. M., & Intille, S. (2019) How intensive longitudinal data can stimulate advances in health behavior maintenance theories and interventions. *Translational Behavioral Medicine* 11(1), 281–286. <https://doi.org/10.1093/tbm/ibz165>
2. Curran, P. J., & Hancock, G. R. (2021) The Challenge of Modeling Co-Developmental Processes over Time. *Child Development Perspectives* 15(2), 67–75. <https://doi.org/10.1111/cdep.12401>
3. Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017) Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal* 25(3), 359–388. <https://doi.org/10.1080/10705511.2017.1406803>
4. McNeish, D., & Hamaker, E. L. (2020) A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods* 25(5), 610–635. <https://doi.org/10.1037/met0000250>
5. Roys, M., Weed, K., Carrigan, M., & MacKillop, J. (2016) Associations between nicotine dependence, anhedonia, urgency and smoking motives. *Addictive Behaviors* 62, 145–151. <https://doi.org/10.1016/j.addbeh.2016.06.002>

6. Dierker, L., Rose, J., Selya, A., Piasecki, T. M., Hedeker, D., & Mermelstein, R. (2015) Depression and nicotine dependence from adolescence to young adulthood. *Addictive Behaviors* 41, 124–128. <https://doi.org/10.1016/j.addbeh.2014.10.004>
7. Asparouhov, T., & Muthén, B. (2018) Latent Variable Centering of Predictors and Mediators in Multilevel and Time-Series Models. *Structural Equation Modeling: A Multidisciplinary Journal* 26(1), 119–142. <https://doi.org/10.1080/10705511.2018.1511375>
8. Muthén, B. O., Muthén, L. K., & Angeles. (1998-2017) *Mplus User's Guide. Eighth Edition* (Muthén & Muthén) 8th ed. Los Angeles, CA.
9. Dou, L., & Hodgson, R. J. W. (1995) Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation. I. *Inverse Problems* 11(5), 1069–1085. <https://doi.org/10.1088/0266-5611/11/5/011>
10. Christopher De Sa, Kunle Olukotun, & Ré, C. (2016) Ensuring Rapid Mixing and Low Bias for Asynchronous Gibbs Sampling. *PubMed* 48, 1567–1576.
11. Jensen, C. S., Kjærulff, U., & Kong, A. (1995) Blocking Gibbs sampling in very large probabilistic expert systems. *International Journal of Human-Computer Studies* 42(6), 647–666. <https://doi.org/10.1006/ijhc.1995.1029>
12. Gelman, A., & Rubin, D. B. (1992) Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
13. Whidden, C., & Matsen, F. A. (2015) Quantifying MCMC Exploration of Phylogenetic Tree Space. *Systematic Biology* 64(3), 472–491. <https://doi.org/10.1093/sysbio/syv006>
14. Fabreti, L. G., & Höhna, S. (2021) Convergence assessment for Bayesian phylogenetic analysis using MCMC simulation. *Methods in Ecology and Evolution* 13, 77–90. <https://doi.org/10.1111/2041-210x.13727>
15. R Core Team. (2022) R: The R Project for Statistical Computing. R-Project.org. <https://www.r-project.org/> (accessed September 2023)
16. Hallquist, M. N., & Wiley, J. F. (2018) MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal* 25(4), 621–638. <https://doi.org/10.1080/10705511.2017.1402334>
17. Wickham, H. (2016) Create Elegant Data Visualisations Using the Grammar of Graphics. Tidyverse.org. <https://ggplot2.tidyverse.org> (accessed September 2023)
18. Asparouhov, T. and Muthén, B. (2010) Bayesian Analysis Using Mplus: Technical Implementation. Mplus Technical Report. <http://www.statmodel.com> (accessed September 2023)
19. Bartlett, M. S. (1937) Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series a - Mathematical and Physical Sciences* 160(901), 268–282. <https://doi.org/10.1098/rspa.1937.0109>
20. Benjamini, Y., & Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
21. Zhou, L., Wang, M., & Zhang, Z. (2021) Intensive Longitudinal Data Analyses With Dynamic Structural Equation Modeling. *Organizational Research Methods* 24(2), 219–250. <https://doi.org/10.1177/1094428119833164>

ABOUT STUDENT AUTHOR

Suryadyuti Baral is currently a senior at UNC - Chapel Hill. He is pursuing a double major in Psychology and Statistics. After graduation he wishes to pursue a PhD in quantitative or computational psychology.

PRESS SUMMARY

Dynamic Structural Equation Models (DSEM) have been widely advertised as a powerful and versatile modeling technique that can shed light on enduring inquiries in the field of psychology. However, our work has unearthed some disconcerting issues in the modeling framework. Turns out DSEM produces estimates that are very unstable and sensitive to arbitrary initializations of its estimation procedure. Its erroneous results can mislead applied researchers to form wrong conclusions from their data. Therefore, it is pivotal that researchers are not left uninformed of these concerns. The study exposes these shortcomings and offers directions for best practice to navigate the deficiencies for anyone who wishes to use DSEM in their work.

Finding the Fixing Number of Johnson Graphs $J(n, k)$ for $k \in \{2, 3\}$

James Della-Giustina

Department of Mathematics, Towson University, Towson, MD

<https://doi.org/10.33697/ajur.2023.097>

Student: jdellagiustina1@gmail.com*

Mentors: Lindsey-Kay Lauderdale: lindseykay.lauderdale@siu.edu, Jay Zimmerman: jzimmerman@towson.edu

ABSTRACT

The graph invariant, aptly named the fixing number, is the smallest number of vertices that, when fixed, eliminate all non-trivial automorphisms (or symmetries) of a graph. Although many graphs have established fixing numbers, Johnson graphs, a family of graphs related to the graph isomorphism problem, have only partially classified fixing numbers. By examining specific orbit sizes of the automorphism group of Johnson graphs and classifying the subsequent remaining subgroups of the automorphism group after iteratively fixing vertices, we provide exact minimal sequences of fixed vertices, in turn establishing the fixing number of infinitely many Johnson graphs.

KEYWORDS

Graph Automorphism Groups; Symmetry Breaking; Fixing Number; Determining Number; Johnson Graphs; Kneser Graphs; Graph Invariants; Permutation Groups; Minimal Sized Bases

INTRODUCTION

Since Euler first posed the Seven Bridges of Königsberg problem in 1736, graph theory has rapidly evolved into a prominent field of research in mathematics. Deceptively simple, graphs provide a powerful mathematical description of the relationships between elements of a system. Graphs may be used to represent physical structures such as organic molecules, electrical circuits, and road systems, in addition to non-physical networks like social groups, data structures, and linguistics.¹⁻⁵

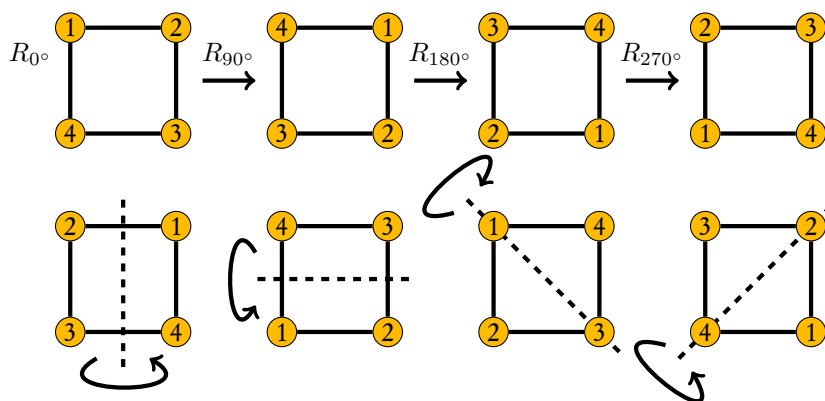


Figure 1. A 2-regular graph on four vertices and four edges has 8 automorphisms, with an automorphism group isomorphic to the dihedral group D_4 .

A graph $\Gamma = (V, E)$ is an unordered double composed of a non-empty vertex set V and an edge set E , consisting of

unordered pairs of vertices from V . A vertex v_i is adjacent to the vertex v_j if and only if there exists an edge $v_i \sim v_j \in E$. The degree of a vertex v is the number of edges containing v , and a k -regular graph is defined by all vertices having equal degree k . Only simple graphs with one connected component are considered in this paper (except for the null graphs in **Proposition 2**), meaning undirected edges with two distinct vertices and at most one edge between vertices.

An automorphism, or symmetry, of a graph Γ is a bijection $\phi : V \rightarrow V$ that preserves the adjacency relations of E such that $\phi(v_i) \sim \phi(v_j) \in E$ if and only if $v_i \sim v_j \in E$. The set of all automorphisms of Γ forms a permutation group on the vertex set V , called the automorphism group of Γ , and is denoted by $\text{Aut}(\Gamma)$.⁶ **Figure 1** presents all eight automorphisms of a 2-regular graph on four vertices and four edges. It should be noted that not all graphs have automorphism groups consisting of different movements. In fact, Erdős showed that the majority of graphs possess only the identity automorphism; such graphs are fittingly named rigid graphs.⁷

We now present a graph invariant that quantifies the complexity of a graph. The fixing number of a graph Γ , denoted $\text{fix}(\Gamma)$, is the minimum number of vertices that, when fixed, eliminate all nontrivial automorphisms of the graph Γ . This minimal sized vertex subset produces a compact representation of the graph’s structure, since any of its automorphisms can be characterized by its action on at least one of these fixed vertices. Also known as symmetry breaking, eliminating graph automorphisms has numerous applications such as characterizing the synthesis of amino acids or assisting robotic systems in determining orientation.^{8,9} This topic has also been investigated under different guises; the determining number or metric dimension of a graph, or, in a purely algebraic view, a minimal sized base of a permutation group acting faithfully on a set.¹⁰

We now state a few results on established fixing numbers for certain graph families. For ease of notation, let $[m, n]$ represent the set $\{m, m + 1, m + 2, \dots, n - 1, n\}$ for $m, n \in \mathbb{Z}$ and $m < n$; the set $\{1, \dots, n\}$ is denoted as $[n]$.

Proposition 1. For a cycle graph C_n with $n \geq 3$, $\text{fix}(C_n) = 2$.

Proof. A cycle graph on n vertices is defined as $C_n = ([n], \{i \sim j \text{ if and only if } j - i \equiv 1 \pmod n \text{ for all } i, j \in [n], i \neq j\})$, with the automorphism group $\text{Aut}(C_n)$ isomorphic to the dihedral group D_n . The 2-regular graph in **Figure 1** is an example of the cycle graph on 4 vertices C_4 . Fixing any single vertex results in an automorphism group isomorphic to \mathbb{Z}_2 , since a reflection across the longest line of symmetry through the fixed vertex and the face of the graph remains. Therefore, fixing any adjacent vertex to the previously fixed vertex will produce a trivial automorphism group. See **Figure 2** for an example. □

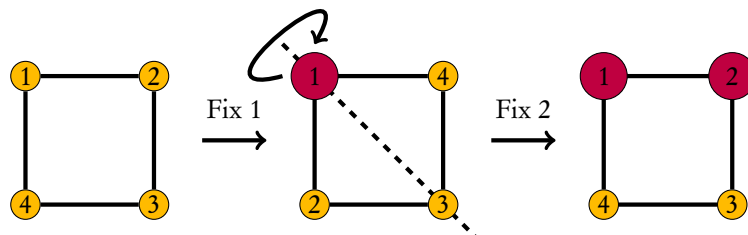


Figure 2. Fixing the vertex labeled 1 leaves a reflection such that the vertices 2 and 4 may be interchanged. Subsequently, fixing the vertex labeled 2 eliminates all nontrivial automorphisms of the graph.

Proposition 2 illustrates that a graph’s connectedness might not impact the determination of the fixing number.

Proposition 2. The complete graph K_n and the null graph N_n on n vertices have fixing number equal to $n - 1$.

Proof. The complete and null graphs are defined as $K_n = ([n], \{i \sim j \text{ for all } i, j \in [n], i \neq j\})$ and $N_n = ([n], \emptyset)$ respectively. In either graph, any vertex can be mapped to any other vertex, and therefore the total number of automorphisms

is $n!$ with both automorphism groups isomorphic to S_n . Fixing one vertex eliminates exactly n automorphisms and leaves an automorphism group isomorphic to S_{n-1} . Therefore, fixing $n - 1$ vertices results in a trivial automorphism group since the ‘free’ unfixed vertex cannot be interchanged with any other vertex, and therefore is also fixed. \square

Johnson Graphs

The focus now shifts to determining the fixing numbers of an infinite sized subset of Johnson graphs, a two parameter graph family denoted as $J(n, k)$.

Definition 1 (^{6, 11}). For positive integers n and k , the Johnson graph $J(n, k)$ for $2 \leq k < \lfloor \frac{n}{2} \rfloor$ is defined as

1. $V(J(n, k))$ consists of unordered k -tuples of $[n]$.
2. The edge $v_i \sim v_j \in E(J(n, k))$ if and only if $|v_i \cap v_j| = k - 1$, i.e., the number of integers in the intersection of the vertex labels equals $k - 1$.

The following properties can be derived from the definition.

- $|V(J(n, k))| = \binom{n}{k}$.
- $\text{Aut}(J(n, k)) \cong S_n$.
- Johnson graphs are *vertex-transitive*, meaning that $\forall v_i, v_j \in V(J(n, k)), \exists \sigma \in \text{Aut}(J(n, k))$ such that $\sigma(v_i) = v_j$.
- The distance between any two vertices $d(v_i, v_j)$ remains constant under any $\sigma \in \text{Aut}(J(n, k))$, implying Johnson graphs are *distance-transitive*.

The combinatorial identity $\binom{n}{k} = \binom{n}{n-k}$ explains the restriction $k \in [2, \frac{n}{2})$ since $J(n, k) \cong J(n, n - k)$ for $k > \frac{n}{2}$. For $k = 1$, the resulting graph is the complete graph K_n defined in **Proposition 2**, of which the fixing number was shown. The second condition of **Definition 1** describes the edge set of $J(n, k)$. However, *generalized* Johnson graphs $J(n, k, i)$ specify that vertices are adjacent if the intersection of integers in vertex labels is equal to $i \in [0, k - 1]$, of which $i = 0$ defines what are known as Kneser graphs.⁶ While the fixing numbers of Kneser graphs have attracted much attention (under the name determining number), the flexibility in constructing edge sets provided by generalized Johnson graphs allows us to classify the fixing number for a significantly larger number of graphs. We provide a few established results on fixing numbers of Kneser graphs; notation has been adjusted for consistency.

Caceres *et al.* found the fixing number for a large subset of Kneser graphs through the following two theorems.

Theorem 1 (Caceres et al.¹²). Let $k, d \in \mathbb{Z}^+$ such that $k \leq d$ and $d > 2$. If $n = \lfloor \frac{d(k+1)}{2} \rfloor + 1$, then

$$\text{fix}(J(n, k, 0)) = d.$$

Theorem 2 (Caceres et al.¹²). Let $k, d \in \mathbb{Z}^+$ such that $3 \leq k + 1 \leq d$. For all $n \in \mathbb{N}$ such that $\lfloor \frac{(d-1)(k+1)}{2} \rfloor < n < \lfloor \frac{d(k+1)}{2} \rfloor$, then

$$\text{fix}(J(n + 1, k, 0)) = d.$$

It should again be emphasized that establishing the fixing number of Johnson, Kneser, or generalized Johnson graphs are equivalent provided that the two parameters n and k agree.

Though beyond the scope of this paper, Babai’s recent work on reducing the graph isomorphism problem to quasi-polynomial time illustrated the importance of Johnson graphs, which he described as "a source of just unspeakable

misery"^{13, 14}. It is not difficult to see why these graphs pose such a problem; they are $k(n - k)$ -regular with distance preserved under any of the $n!$ automorphisms. The performance improvement of his algorithm stems from the recognition of whether a graph contains a Johnson graph as a subgraph and employing the appropriate methods upon that decision. Babai’s own words reinforce the motivation for our work: “In fact, breaking regularity is one of the key tools in the design of algorithms for graph isomorphism; the graph isomorphism problem has therefore been one of the strongest motivators of the study of all sorts of ‘resolving/discriminating sets’, and perhaps the only deep motivator of the study of those in contexts where no group is present.”¹⁰

Fixing Johnson Graphs

In this section, we establish the fixing number of the Johnson graphs $J(n, k)$ for $n > 2k$ and $k \in \{2, 3\}$. Throughout, we assume for a finite graph $\Gamma = J(n, k)$ the following action of $\text{Aut}(\Gamma)$ on $V(\Gamma)$, which is the standard action of S_n acting on k -subsets.

Definition 2. Let Γ be a graph such that $\text{Aut}(\Gamma) \cong S_n$ acts transitively on the unordered k -subsets of $[n]$ such that for all $\sigma \in S_n$ and $\forall \{\ell_1, \ell_2, \dots, \ell_k\} \in V(\Gamma)$,

$$\sigma \cdot \{\ell_1, \ell_2, \dots, \ell_k\} = \{\sigma(\ell_1), \sigma(\ell_2), \dots, \sigma(\ell_k)\}.$$

Since Johnson graphs are vertex-transitive, there exists only one orbit in $V(\Gamma)$ of size $\binom{n}{k}$.

Theorem 3. If $\Gamma = J(n, 2)$ where $n \geq 5$, then $\text{fix}(\Gamma) = \lfloor \frac{2n}{3} \rfloor$.

Proof. Let $\text{Aut}(\Gamma) \cong S_n$ act transitively on the 2-subsets of $V(\Gamma)$ by the action specified in **Definition 2**. Consider fixing the vertex with integer label $\{1, 2\}$ so that it may not be permuted among any other vertices in $V(\Gamma)$. The remaining non-trivial elements of the automorphism group have exactly two forms

$$(1, 2)\tau \quad \text{or} \quad (1)(2)\tau \tag{1}$$

where τ is any permutation of $[3, n]$. The transposition $(1, 2)$ remains since $\{1, 2\} \equiv \{2, 1\}$, *i.e.*, the vertex is fixed, but the integers 1 and 2 may still permute in the fixed vertex label when the automorphism group acts on $V(\Gamma)$. Therefore, fixing the vertex with integer label $\{1, 2\}$ results in a subgroup of automorphisms isomorphic to $S_{n-2} \times \mathbb{Z}_2$.

Now, fix another element of $V(\Gamma)$, the vertex with integer label $\{2, 3\}$. Every element that had the form $(1, 2)\tau$ is now eliminated, since the integers 1 and 2 are no longer able to permute without also moving the now fixed vertex $\{2, 3\}$. In this sense, the approach can be viewed as finding the optimal method of fixing integers $[n - 1]$ in the vertex labels of elements in $V(\Gamma)$ by correctly choosing the appropriately labeled vertices. A crucial observation of this work is the attempt to only fix the integers $[n]$ if necessary, otherwise fixing $[n - 1]$ achieves the desired result and may result in a smaller fixing number. The remaining non-trivial elements of the automorphism group are represented as

$$(1)(2)(3)\rho$$

where ρ is any permutation of the integers $[4, n]$. An algorithmic approach is to continue fixing vertices with integer labels in a sequence of at most

$$\{1, 2\}, \{2, 3\}, \{4, 5\}, \{5, 6\}, \{7, 8\}, \{8, 9\}, \dots, \{n - 2, n - 1\}, \{n - 1, n\}. \tag{2}$$

Begin by examining the smallest unfixed consecutive integers x, y , and z in the vertex labels of $V(\Gamma)$. At each step in the algorithm, if the vertex label $\{x, y\}$ consists of integers of which neither has been fixed in a previous step, then the automorphism group that remains when fixing $\{x, y\}$ is isomorphic to $S_{n-y} \times \mathbb{Z}_2$. However, if y has already appeared in a

previously fixed vertex label, then by fixing the next vertex label $\{y, z\}$, the resulting automorphism group will be isomorphic to S_{n-z} . We deviate from this algorithm only if $n \equiv 2 \pmod 3$, at which the last 3 fixed vertices have the form

$$\{x, y\}, \{y, z\}, \{z, z + 1\}. \tag{3}$$

By fixing pairs of vertices with consecutive integer labels $\{x, y\}$ and $\{y, z\}$, the resulting automorphism group is isomorphic to S_{n-3} since the integers x, y , and z in the vertex labels of $V(\Gamma)$ are fixed.

To show the upper bound of $\text{fix}(\Gamma) \leq \lfloor \frac{2n}{3} \rfloor$, write $n = 3q + r$ for $q = \lfloor \frac{n}{3} \rfloor$ and $r \equiv n \pmod 3$. By noting that each pair of fixed vertices of the form $\{x, y\}$ and $\{y, z\}$ fixes the three integers x, y , and z in the vertex labels of $V(\Gamma)$, we consider two cases.

- If $r = 0$, a fixing set F can be given as

$$F = \bigcup_{i=1}^q \{\{3i - 2, 3i - 1\}, \{3i - 1, 3i\}\} \tag{4}$$

resulting in a fixing set of size $|F| = 2q$. If $r = 1$, then F fixes the integers $[n - 1]$ in the vertex labels of $V(\Gamma)$, stabilizing the integer n .

- If $r = 2$, then the last three fixed vertex labels are defined in (3). Therefore, fixing $2q$ vertices given by (4) and one additional vertex yields

$$F' = F \cup \{\{3q, 3q + 1\}\}.$$

In either case, we have that $|F| = |F'| = \lfloor \frac{2n}{3} \rfloor$ which shows $\text{fix}(\Gamma) \leq \lfloor \frac{2n}{3} \rfloor$. To show $\text{fix}(\Gamma) \geq \lfloor \frac{2n}{3} \rfloor$, we build upon a combinatorial argument that appeared in Tracey Maund’s unpublished dissertation on base sizes of permutation groups.¹⁵ For ease of notation, define $f = \text{fix}(\Gamma)$ for a fixing set F^* . Clearly, $n = \sum_{i=0}^f n_i$ where n_i is the number of occurrences that an integer is present in i vertices of the fixing set. Similarly, $f \cdot k$ is the total number of places that an integer can occupy in the vertices of a fixing set, since there are f vertices each of size k . So,

$$\begin{aligned} f \cdot k &= \sum_{i=0}^f i \cdot n_i \\ &= 0 \cdot n_0 + 1 \cdot n_1 + 2 \cdot n_2 + \dots + f \cdot n_f \\ &= 2 \cdot (n_1 + n_2 + \dots + n_f) - n_1 + [n_3 + 2 \cdot n_4 + \dots + (f - 2) \cdot n_f] \\ &= 2 \cdot (n - n_0) - n_1 + [n_3 + 2 \cdot n_4 + \dots + (f - 2) \cdot n_f] \\ &\geq 2(n - n_0) - n_1. \end{aligned}$$

Let $n_0 \leq 1$ since $n_0 \geq 2$ implies the transposition $(1, 2) \in \text{Aut}(\Gamma)$, contradicting that F^* is a fixing set. Next, note that $n_i \leq f$, since no integer can appear in more than f vertices. Then

$$\begin{aligned} f \cdot k &\geq 2(n - n_0) - n_1 \geq 2(n - 1) - f \\ f &\geq \frac{2(n - 1)}{k + 1}. \end{aligned}$$

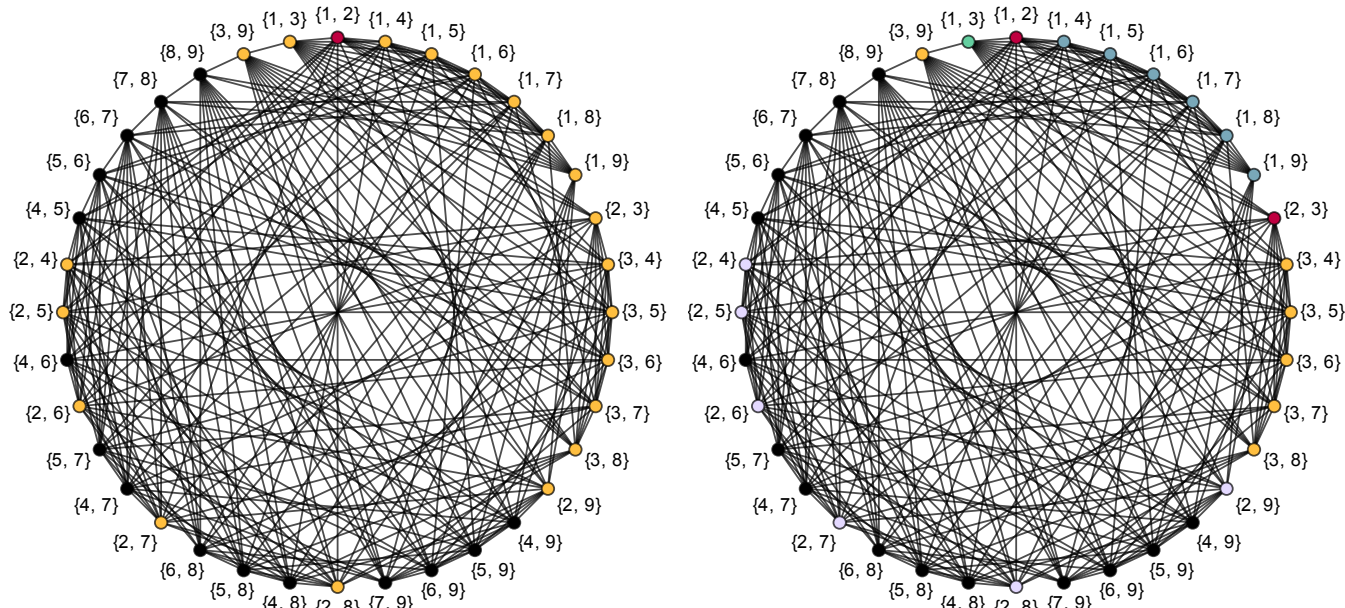
Since $f \in \mathbb{N}$, then $f \geq \lceil \frac{2(n-1)}{k+1} \rceil$. In general, for $s, t \in \mathbb{Z}$, $\lceil \frac{s}{t+1} \rceil = \lfloor \frac{s+t}{t+1} \rfloor$, and so $\lceil \frac{2(n-1)}{k+1} \rceil = \lfloor \frac{2(n-1)+k}{k+1} \rfloor$. For $k = 2$, this reduces down to

$$\begin{aligned} f &\geq \left\lfloor \frac{2(n - 1) + 2}{2 + 1} \right\rfloor \\ \text{fix}(\Gamma) &\geq \left\lfloor \frac{2n}{3} \right\rfloor. \end{aligned}$$

Therefore, $\text{fix}(\Gamma) = \lfloor \frac{2n}{3} \rfloor$. □

We now provide an example in order to classify the subgroups of the automorphism group that remain when iteratively fixing vertices and the algorithmic approach for choosing specific vertex labels.

Example 1. Consider the Johnson graph $J(9, 2)$ which has unordered doubles of $[1, 9]$ as vertices and edges $v_i \sim v_j$ if the respective vertex labels share exactly one integer entry, seen in **Figure 3**. First, fix the vertex with integer label $\{1, 2\}$. All remaining non-trivial elements in the automorphism group have the form ψ or $(1, 2)\psi$ where ψ is any permutation of the integers $[3, 9]$, indicating a subgroup of automorphisms isomorphic to $\mathbb{Z}_2 \times S_7$. Vertices with labels containing only the integers $[3, 9]$ can freely permute, but vertices with integer labels $\{i, j\}$ for $i \in [1, 2]$ and $j \in [3, 9]$ now form a separate orbit of size 14, illustrated by the gold vertices seen in **Figure 3a**. Proceed by fixing the vertex with the integer label $\{2, 3\}$. Ev-



(a) Fixing the vertex $\{1, 2\}$ reduces the total number of automorphisms from 362,880 to 10,080.

(b) Fixing the red vertices with integer labels $\{1, 2\}$ & $\{2, 3\}$ consequently fixes the green vertex with label $\{1, 3\}$.

Figure 3. Fixing two vertices of $J(9, 2)$ with integer labels $\{1, 2\}$ & $\{2, 3\}$ reduces the size of the automorphism group from 362, 880 to 720, a 99.8% decrease.

ery remaining non-trivial element ν of the automorphism group consists of permutations of $[4, 9]$, since the transposition $(1, 2)$ can no longer act on $V(\Gamma)$ without permuting the vertex labeled $\{2, 3\}$. Although only two fixed vertices have been chosen, now the vertex with label $\{1, 3\}$ is also fixed as a result. Any of the gold, blue, or pink vertices seen in **Figure 3b** with integer labels $\{i, j\}$ for $i \in [1, 3]$ and $j \in [4, 9]$ may only permute among vertices of the same color (*i.e.*, sharing the same i value; black vertices may still be mapped freely), indicating the remaining automorphism group is isomorphic to S_6 .

Following the sequence of fixed vertices given by (2), fixing vertices with integer labels

$$\{1, 2\}, \{2, 3\}, \{4, 5\}, \{5, 6\}, \{7, 8\}, \{8, 9\},$$

yields the associated chain of subgroups of S_9

$$S_7 \times \mathbb{Z}_2 \geq S_6 \geq S_5 \times \mathbb{Z}_2 \geq S_4 \geq S_3 \times \mathbb{Z}_2 \geq \mathbb{Z}_2 \geq \{e\}.$$

Therefore, the fixing number of $J(9, 2)$ is equal to $\lfloor \frac{2(9)}{3} \rfloor = 6$, since any subset of 5 fixed vertices will leave at least one remaining transposition in the automorphism group.

We now utilize much of the same machinery from the proof of **Theorem 3** to establish the fixing numbers of Johnson graphs $J(n, 3)$ for $n \geq 7$.

Theorem 4. If $\Gamma = J(n, 3)$ where $n \geq 7$, then $\text{fix}(\Gamma) = \lfloor \frac{n}{2} \rfloor$.

Proof. Let $\text{Aut}(\Gamma) \cong S_n$ act transitively on the 3-subsets of $V(\Gamma)$ by the action stated in **Definition 2** and consider fixing the vertex with integer label $\{1, 2, 3\}$. The remaining non-trivial elements of the automorphism group have the following 6 forms,

$$(1, 2)\tau \quad (1, 3)\tau \quad (2, 3)\tau \quad (1, 2, 3)\tau \quad (1, 3, 2)\tau \quad \tau \tag{5}$$

where τ is any permutation of $[4, n]$. Therefore, the subgroup of automorphisms consisting of all forms in (5) and the identity automorphism is isomorphic to $S_3 \times S_{n-3}$.

Fix another vertex with the integer label $\{3, 4, 5\}$. Every element of S_n that had a cycle containing the integer 3 is now eliminated, since each of the previous two fixed vertex labels contained 3. The remaining non-trivial elements in the automorphism group each have one of the following 7 forms,

$$(1, 2) \quad (4, 5) \quad (1, 2)(4, 5) \quad \rho \quad (1, 2)\rho \quad (4, 5)\rho \quad (1, 2)(4, 5)\rho$$

for any permutation ρ of the integers $[6, n]$. Subsequently, the remaining subgroup of the automorphism group is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times S_{n-5}$. Next, fix the vertex with integer label $\{5, 6, 7\}$, in turn fixing the integers $[3, 5]$ and resulting in an automorphism group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times S_{n-7}$ (the transpositions $(1, 2)$ and $(6, 7)$ remain). Continue fixing vertices of consecutive integers that have the first and last integers appearing in some other fixed vertex label. However, note that the integer 1 needs to appear in the last fixed vertex label chosen to eliminate the remaining transposition $(1, 2)$.

The general approach is to proceed by fixing vertices with integer labels in a sequence of at most,

$$\{1, 2, 3\}, \{3, 4, 5\}, \{5, 6, 7\}, \{7, 8, 9\}, \dots, \{n-5, n-4, n-3\}, \{n-3, n-2, n-1\}, \{n-1, n, 1\}.$$

Start by examining the smallest unfixed consecutive integers s, t , and u in the vertex labels of $V(\Gamma)$ (note that s has appeared in a previously fixed vertex label). Fixing the vertex with label $\{s, t, u\}$ results in an automorphism group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times S_{n-u}$. The next fixed vertex will have an integer label of the form $\{u, v, x\}$ for consecutive integers u, v , and x and produces an automorphism group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times S_{n-x}$, since the transpositions $(1, 2)$ and (v, x) remain but the integer u is fixed. Fixing another vertex with integer label $\{x, y, z\}$ eliminates one transposition (v, x) but introduces the transposition (y, z) , and so the automorphism group is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times S_{n-z}$. Similar to **Theorem 3**, observe that fixing $3r$ vertices fixes $6r - 3$ integers in the vertex labels of $V(\Gamma)$.

Now, write $n = 2q + s$ for $q = \lfloor \frac{n}{2} \rfloor$ and $s \equiv n \pmod 2$. If the parity of n is even, then the last vertex to fix will have the form $\{n-1, n, 1\}$ and a fixing set of size equal to $\frac{n}{2}$. For n odd, the last vertex will have the integer label $\{n-2, n-1, 1\}$, and consequently the fixing set will be of size $\frac{n-1}{2}$. In either case, the fixing set F can be given by

$$F = \{\{2q-1, 2q, 1\}\} \cup \bigcup_{i=1}^{q-1} \{2i-1, 2i, 2i+1\},$$

and so $\text{fix}(\Gamma) \leq \lfloor \frac{n}{2} \rfloor$.

To show the opposite inequality, recall $n = \sum_{i=0}^f n_i$ where n_i is the number of integers that appear i times in the vertices of a fixing set. Since $f \cdot k$ is the number of entries integers can occupy in the vertices of the fixing set then, similar to **Theorem 3**, we have $f \cdot k \geq 2(n - n_0) - n_1$. Since $n_i \leq f$ we choose $n_1 = f$ to be of maximal value, but now consider two cases based upon $n_0 \equiv n \pmod 2$.

- For $n_0 = 0$,

$$\begin{aligned}
 f \cdot k &\geq 2n - f \\
 &= \left\lceil \frac{2n}{k+1} \right\rceil \\
 &= \left\lceil \frac{2n}{4} \right\rceil \\
 &= \left\lceil \frac{n}{2} \right\rceil.
 \end{aligned}$$

Since n is even, then $\lceil \frac{n}{2} \rceil = \lfloor \frac{n}{2} \rfloor$.

- For $n_0 = 1$,

$$\begin{aligned}
 f \cdot k &\geq 2(n-1) - f \\
 &= \left\lceil \frac{2(n-1)}{k+1} \right\rceil \\
 &= \left\lceil \frac{2(n-1)}{4} \right\rceil \\
 &= \left\lceil \frac{n-1}{2} \right\rceil.
 \end{aligned}$$

Given that n is odd, then $\lceil \frac{n-1}{2} \rceil = \lfloor \frac{n-1}{2} \rfloor$ which is equivalent to $\lfloor \frac{n}{2} \rfloor$.

In either case, we have that $\text{fix}(\Gamma) \geq \lfloor \frac{n}{2} \rfloor$ and therefore $\text{fix}(\Gamma) = \lfloor \frac{n}{2} \rfloor$. □

CONCLUSIONS

In this work, the fixing number for an infinite sized subset of Johnson graphs was established by studying the action of the automorphism group on the vertex set, methodically choosing which vertices to fix by the associated integer labels, and observing the resulting automorphism groups. The fixing numbers calculated in **Theorems 3 & 4** unify the two separate results established by Caceres *et al.* in **Theorems 1 & 2** for specific values of k . Future work includes publishing fixing number results for an even greater subset of Johnson graphs for $k \geq 4$ and exploring the viability of applying similar number schemes in a cryptological system.

ACKNOWLEDGEMENTS

A sincere thanks to the anonymous reviewer, whose invaluable advice helped streamline and simplify many of the proofs. The author also thanks the NSA (Grant #H98230-20-1-0005) for funding the Research Experience for Undergraduates (REU) that made this work possible along with his team members, Ummul Aymen, Jason Riley Jr., and Erin Stales. In addition to his excellent research advisors Dr. Lindsey-Kay Lauderdale & Dr. Jay Zimmerman, he would also like to thank the numerous Towson University faculty (Drs. Kolesnikov, Nganguia, Cornwell, Kumchev, Gluck, McNew, & Gunning) for their priceless advice and perspectives in mathematics and everything else. Finally, a deep gratitude to Tracey Maund, who generously provided a copy of her unpublished thesis and graciously took the time to explain specific concepts through a detailed correspondence.

REFERENCES

1. Kirsch, R. A., and Ray, L. C. (1957) Finding chemical records by digital computers, *Science* 126, 814–819. <https://doi.org/10.1126/science.126.3278.814>
2. Vrbik, J. (2022) Solving electrical circuits via graph theory, *Applied Mathematics* 13, 77–86. <https://doi.org/10.4236/am.2022.131007>
3. Unsalan, C., and Sirmacek, B. (2012) Road network detection using probabilistic and graph theoretical methods, *IEEE Transactions on Geoscience and Remote Sensing* 50, 4441–4453. <https://doi.org/10.1109/TGRS.2012.2190078>

4. Barnes, J. A. (1969) Graph theory and social networks: a technical comment on connectedness and connectivity, *Sociology* 3, 215–232. <https://doi.org/10.1177/003803856900300205>
5. Kuhlmann, M., and Oepen, S. (2016) Towards a catalogue of linguistic graph banks, *Computational Linguistics* 42, 819–827. https://doi.org/10.1162/COLI_a00268
6. Godsil, C., and Royle, G. (2001) Graphs, in *Algebraic Graph Theory* (Axler, S., Gehring, F.W., Ribet, K.A. Ed.) 1st ed., 4–10, Graduate Texts in Mathematics Springer-Verlag, New York. 10.1007/978-1-4613-0163-9
7. Erdős, P., and Rényi, A. (1963) Asymmetric graphs, *Acta. Math. Hungarica.* 14, 295–315. <https://doi.org/10.1007/BF01895716>
8. Antoneli, F., and Forger, M. (2010) Symmetry breaking in the genetic code: finite groups, *Math. Comput. Model.* 53, Iss. 7–8, 1469–1488. <https://doi.org/10.1016/j.mcm.2010.03.050>
9. Lynch, K. M. (2001) Determining the orientation of a painted sphere from a single image: a graph coloring problem, *CiteSeerX*, <https://tinyurl.com/d5ey9wfb>
10. Bailey, R., and Cameron, P. (2011) Base size, metric dimension and other invariants of groups and graphs, *Bull. London Math. Soc.* 43, 209–242. <https://doi.org/10.1112/blms/bdq096>
11. Brouwer, A. E., Cohen, A. M., and Neumaier, A. (1989) Families of graphs with classical parameters, in *Distance-Regular Graphs* 1st ed., 255–256, *Ergeb. Math. Grenzgeb* vol 18. Springer, Berlin, Heidelberg. 10.1007/978-3-642-74341-2
12. Cáceres, J., Garijo, D., González, A., Márquez, A., and Luz Puertas, M. (2013) The determining number of Kneser graphs, *Discret. Math Theor. Comput. Sci.* 15, 1–14. <https://doi.org/10.46298/dmtcs.634>
13. Klarreich, Errica (2015) Landmark algorithm breaks 30-year impasse, *Quanta Magazine*, <https://www.quantamagazine.org/algorithm-solves-graph-isomorphism-in-record-time-20151214> (accessed July 2021)
14. Babai, László (2016) Graph isomorphism in quasipolynomial time, *Proc. Annu. ACM Symp. Theory Comput.* 48, 648–697, <https://doi.org/10.1145/2897518.2897542>
15. Maund, Tracey (1989) Bases for permutation groups, *Dissertation, University of Oxford*

ABOUT THE STUDENT AUTHOR

James Della-Giustina graduated from Towson University with a B.S. in Spring '22 and is slated to earn an M.S in Fall '23; both in applied mathematics. After graduation, he hopes to gain a few years of industry experience before pursuing a Ph.D. focused in algebra and discrete structures.

PRESS SUMMARY

We study a concept called the fixing number, which is the smallest number of vertices needed to eliminate all symmetries (mappings consisting of rotations, reflections, etc.) of a mathematical object called a graph. While many graphs have known fixing numbers, the fixing numbers of Johnson graphs - a family of graphs relevant to the study of graph isomorphism - are only partially understood. By iteratively fixing vertices and analyzing the sizes of resulting subgroups within the automorphism group, we are able to determine the fixing number for infinitely many Johnson graphs. Our findings have important implications for understanding the structure of these graphs and could have applications in fields such as computer science and cryptography.

Anxiety as a Predictor of Movement During a Math Task in Elementary School Children

Gabriella Snetkov^{a*}, Anna E. Youngkin^b, Megan E. Lutz^c, Analia Marzoratti^a, & Tanya M. Evans^{a,b}

^aSchool of Education and Human Development, University of Virginia, Charlottesville, VA

^bSchool of Medicine, University of Virginia, Charlottesville, VA

^cBoston Children's Hospital, Boston, MA

<https://doi.org/10.33697/ajur.2023.098>

Students: gabby.snetkov@gmail.com^{*}, gyb9cr@virginia.edu

Mentor: tanya@virginia.edu

ABSTRACT

Anxiety can come in various forms; general anxiety is characterized by feelings of excessive anxiety impacting multiple domains of everyday life, whereas anxiety may also manifest within a single domain (*i.e.*, math anxiety). Those that present with anxiety who also have Autism Spectrum Disorder often exhibit ritualized movement as a coping mechanism for their anxiety. However, the relationship between anxiety and movement has not yet been explored within typically developing children. It is also unclear if the form of anxiety impacts the degree to which children move. In this study, we used Motion Energy Analysis to quantify the relationship between both general and math anxiety and movement during a math flashcards task. Multivariate linear regression analysis was used to determine possible relationships between anxiety and movement. General anxiety significantly contributed to models predicting movement during the math flashcards task, while math anxiety did not. Our results suggest that movement could be an indicator of general anxiety in elementary aged children.

KEYWORDS

General Anxiety; Math Anxiety; Elementary; Academics; Motion Energy Analysis; Movement; Ritualized Movement; Mathematics

INTRODUCTION

Anxiety is a psychiatric disorder impacting approximately 12.5% of the adult population in the United States. Additionally, 9.4% of children ages 3-17 years will be diagnosed with anxiety, with the proportion of children diagnosed with anxiety growing substantially over the last decade.¹⁻³ In addition to affecting day-to-day experiences, anxiety can have detrimental effects in academic and educational settings.⁴ Anxiety in students has been associated with reduced thriving and academic achievement in students at multiple ages.⁵⁻⁶ Importantly, the negative impacts of general anxiety are not subject specific, having measurable negative impact across academic subject areas.⁷ Additionally, student anxiety scores have been shown to increase as students' progress from elementary to middle to high school, which furthers the impact of anxiety on academic performance.⁸ The growing impact of anxiety on students motivates the need to determine how anxiety might manifest in students of all ages.

In some cases, anxiety can be associated with a specific academic subject, such as math-specific anxiety, which affects approximately 17% of the U.S. student population. Importantly, unlike general anxiety which has measurable impacts across academic subject areas⁷, math anxiety negatively influences math-specific skills, grades in math classes, and general attitudes towards math.^{4, 9, 10} Math anxiety is also known to impact student choices and behavior, as students with math anxiety tend to avoid taking math-related classes altogether.¹¹ While both general and math anxiety have similar impacts on academic performance, the relationship between the anxiety types is complicated. While those with math anxiety are much more likely to also exhibit general anxiety symptoms,⁹ presence of general anxiety by itself has not been shown to predict math performance.¹² As a result, while individuals with math anxiety may be more likely to have generalized anxiety, general anxiety alone often does not lead to anxiety-related math deficits. This complicated relationship implores the need for research into how these different anxiety types manifest within students and impact their behavior.

One behavior that may be differentially impacted by generalized vs. subject-specific anxiety is soothing behavior, which is the variety of behaviors individuals dealing with anxiety turn to in order to soothe their anxious emotions. Ritualized movements, often referred to as stimming, are known to strongly correlate with anxiety symptoms within individuals diagnosed with Autism Spectrum Disorder (ASD).¹³ Additionally, placing individuals with ASD into scenarios that often trigger anxiety, such as public speaking, has been shown to cause ritualized movement behavior.¹⁴ However, it is unclear whether ritualized movements are

uncontrolled responses to anxiety, or are instead self-regulatory behavior to sooth anxiety symptoms.¹⁴⁻¹⁶ In fact, individuals with ASD of various ages report that ritualized movements provide emotional regulation support, despite the behavior not consistently being deemed socially acceptable.¹⁷⁻²⁰

The current body of knowledge surrounding the relationship between movement and anxiety is mostly centered around individuals with ASD. As a result, it is unknown if movement might also have important relationships with anxiety in typically developing children, with anxiety leading to increased movement. Additionally, no current literature has explored whether various anxiety types have differing predictive contributions to movement. The current study aims to fill these gaps in understanding. We quantified movement in typically developing children ages 6-11 years with motion energy analysis (MEA) while the subjects completed a math flashcards task. Our results examine the independent and overlapping effects of general and math anxiety on movement to determine whether anxiety contributes to movement and explore whether one form of anxiety contributes to movement more than the other. Overall, this research aims to broaden the understanding of how anxiety impacts behavior during elementary school age.

METHODS AND PROCEDURES

Sample

Parent-child dyads were both indirectly and directly recruited from the greater Charlottesville-Albemarle County Area for study participation. Families either responded to flyers posted around the community or were contacted by research personnel using a University of Virginia-managed database of families who had previously noted interest in research participation. Parents completed an eligibility questionnaire which they populated with family demographic information and child medical history. Inclusion criteria included child age between 6-11 years, indication of the child's ability to communicate verbally, and no diagnosis of a neurodevelopmental disorder. Parents were required to be 18 years of age or older. Parents provided written informed consent while the child provided verbal assent unless over the age of seven, in which case written assent was obtained. Dyads were excluded from the final sample if the child's Intelligence Quotient (IQ) was less than 85, the child could not complete the experimental session due to behavioral issues or discomfort with equipment, or if there was not a complete dataset for the dyad (*e.g.*, missing scores for a cognitive assessment subset). The final sample consisted of dyads of 15 typically developing children (aged 9.33 ± 1.64 years, 8 male) and one parent per child (4 male parents).

Procedure

All recruitment and protocol procedures were approved by the University of Virginia Social and Behavioral Sciences Institutional Review Board (SBS-2174). The data utilized here is a subset of a larger study that included children with Autism Spectrum Disorder in addition to the typically developing sample. At the Sheila Johnson Center for Clinical Services at the University of Virginia, participants completed two separate sessions as part of this study: a cognitive assessments session and an experimental task session which occurred on two separate days.

During cognitive assessment sessions, a trained clinician administered a battery of assessments to pediatric participants while parents waited in a separate room. The Woodcock-Johnson IV (WJ-IV) was used to assess children's cognitive abilities and the Differential Ability Scales Version 2 (DAS-II) was used to assess child IQ (*e.g.*, quantitative, verbal, and nonverbal reasoning). Of interest to this analysis is the math facts fluency subtest of the WJ-IV. In this subtest, participants are presented with a series of simple addition, subtraction, and multiplication questions. Participants are instructed to complete as many facts as possible within the three-minute time limit, and their performance is scored based on the number of correct answers and a qualitative observation of their rate of performance compared to their peers. To assess children's math anxiety, the Single Item Math Anxiety (SIMA) scale was administered.¹⁰ In this assessment, children were asked the following question: "On a scale of one to 10, how math anxious are you?" Math anxiety was described to child participants as feelings of anxiety or stress that occur when learning or doing math, with the scale depicting a score of one as "not anxious" and a 10 as "very anxious." Children also provided an analogous general anxiety measure by answering the question: "On a scale of one to 10, how anxious are you?" using the same scale. Children also completed a math excitement questionnaire. Participants were presented with a scale of five drawings of faces depicting varying levels of excitement, labeled "not at all excited", "not very excited", "just a little excited", "very excited", or "very, very excited". Each drawing was given a score between zero and four, with higher scores indicating greater excitement. Participants were instructed to mark a line through the face that best described their response to the following questions: "How excited do you feel when your teacher talks about numbers and math," "How excited do you feel when you have to do number and math work at school," and "How excited do you feel when you are asked to practice numbers and math at home?". The sum of the scores for each question served as total math excitement score, with possible values ranging from zero to 12.

During experimental sessions, children and their parents performed a variety of tasks designed to elicit varying types of interaction while sitting across from each other at a table. This analysis focuses specifically on a math flashcards task, where parents were instructed to quiz their child using arithmetic flashcards. The task mandated that parents quiz the child (not vice versa) and that

they proceeded through the cards in a specified order. Parents were allowed to provide as much or as little help as they deemed appropriate. Parents were provided with three flashcard decks, each consisting of 25 cards, and instructed to proceed through them in order of ascending difficulty. Stacks one and two contained single-digit and double-digit addition and subtraction problems, while stack three contained single-digit multiplication problems. The cards of each deck were numbered in a way such that all dyads answered the same cards in the same order. The task lasted approximately five minutes, and participants were instructed to complete as many cards as possible in that time frame. If a child finished answering all three decks of flashcards within the time limit, the participants were instructed to start again from the beginning to ensure five minutes of recorded data for all dyads.

Motion Energy Analysis (MEA)

Experimental sessions were recorded on video and analyzed post-hoc for total amount of child movement during the task using the Motion Energy Analysis (MEA) software. MEA is a software that quantifies the degree of change in pixilation between frames in a user-defined area of interest, which can be used as a proxy for movement within a specific region of interest.²¹ Video recordings provided three angles of the subjects during the task. For this analysis, a front-facing view of the child participant was used, with the region of interest defined to capture as much of the child’s full-body motion as possible. An MEA time series was produced for each child participant for this specific math flashcards task condition. The threshold for movement detection was changed from the default value of “20” to “seven.” This value was identified during initial processing to be the most appropriate threshold for the quality of our recorded videos in distinguishing noise from legitimate, finer movements in the video. This identification was accomplished using the “frame-differenced movie” preview of the MEA software, following recommendations of choosing a threshold that yields a clear white outline around the participant of interest and minimal noise elsewhere in the video.

Statistical Analysis

Data was uploaded into SPSS for statistical analysis. Multivariate linear regression models were used to determine which components of anxiety produced better predictive models of child movement. Median child MEA scores were used as the dependent variable. Median values were used as opposed to mean values to better account for periods of no movement at the very beginning and the very end of the task. Independent variables of interest included: age, sex, IQ, WJ-IV math facts fluency, and reported excitement related to math. Three separate models were run: (1) general anxiety scores were added as an additional independent variable, (2) math-specific anxiety scores were added as an independent variable, and (3) both general and math anxiety scores were added as independent variables. The variables were removed from the model using a backwards approach with criteria of probability of $F > 0.1$, meaning any variables in the model that have a p-value greater than 0.1 were removed one at a time until all variables contributed to the predictive power of the model at a threshold of $p \leq 0.1$. As a result, only variables with relevant predictive effects were included in the final models. We determined the most successful model to be that which had the highest adjusted R^2 and had a p-value of < 0.05 .

RESULTS

Group Characteristics

The final study sample consisted of seven female and eight male children with an average age of 9.33 (SD=1.64) years. The children had an average full-scale IQ of 116.4 (SD=12.81) and math facts fluency score of 95.47 (SD=9.478). General anxiety scores of children ranged from one to nine, with an average of 4.4 (SD=3.22). Math anxiety scores of children ranged from one to eight with an average of 4.8 (SD=2.21). There was no significant difference between male and female subjects’ scores on general anxiety or math anxiety questionnaires, or between math facts fluency score or IQ scores as determined by unpaired T-Tests. Median MEA value across all subjects was 397 (SD=352) pixels/frame. Complete characteristics for all variables are included in Table 1.

	Mean	SD	Range	Skew	Kurtosis	n
Age (years)	9.33	1.64	6.3-11.7	-0.481178	-0.774209	15
IQ	116.4	12.81	101-136	0.674416	-0.708996	15
Math Facts Fluency	95.47	9.748	75-108	-0.874916	0.644223	15
General Anxiety	4.4	3.22	1-9	0.217035	-1.868782	15
Math Anxiety	4.8	2.21	1-8	0.428603	-0.72397	15
Math Excitement	6.067	2.79	3-12	0.536207	0.03595	15
Median MEA (pixels)	397.13	352.015	8-1236	1.161488	0.973373	15

Table 1. Characteristics of Included Variables

Anxiety Contributions to Movement During Task

Math Anxiety Model

Linear Regression Models were used to examine the relationship between math anxiety and child movement. Predictors included in the models were age, sex, IQ, math facts fluency score, math excitement and math anxiety. None of these models were significant in predicting movement during the math task (Supplementary Table 1).

General Anxiety Model

To examine the relationship between movement and general anxiety, the same predictors above were used, but math anxiety was replaced with general anxiety. The best-fitting linear model included general anxiety, age, math excitement, and IQ as predictor variables for child movement (adjusted R² = 0.539, p = 0.017) (Table 2). Within this model, unit increases in age (p = 0.039), math excitement (p = 0.049), and general anxiety (p = 0.005) were found to have significant impacts on child movement. Age was negatively associated with movement (B=-128.334), whereas math excitement (B=77.507) and general anxiety (B=69.001) were positively associated with movement. IQ was also positively associated with movement, although not significantly (B=6.199, p=0.278) (Table 4).

Model	R	R Square	Adjusted R Square	F	p-value
1	.825 ^a	.680	.440	2.836	.087
2	.823 ^b	.677	.498	3.775	.040
3	.819^c	.670	.539	5.085	.017
4	.792 ^d	.627	.525	6.165	.010

- a. Predictors: (Constant), Math Facts Fluency, General Anx, Age, Math Excitement, Sex, IQ
- b. Predictors: (Constant), Math Facts Fluency, General Anx, Age, Math Excitement, IQ
- c. Predictors: (Constant), General Anx, Age, Math Excitement, IQ
- d. Predictors: (Constant), General Anx, Age, Math Excitement

Table 2. Modeling movement using general anxiety

Math Anxiety and General Anxiety Model

We then determined whether movement could be better modeled using a combination of math and general anxiety using the same modeling structure and predictors but including both math and general anxiety. The best-fitting linear model for movement included general anxiety, age, math excitement, and math anxiety as predictor variables (adjusted R² = 0.640, p = 0.005) (Table 3). Within this model, unit increases in general anxiety (p = 0.007), math excitement (p = 0.007), and age (p = 0.008) were again found to significantly impact child movement. General anxiety and math excitement were again both positively associated with movement and age was negatively associated, reflecting the findings from the general anxiety model. Math anxiety (B = 41.703, p = 0.141) did not have a significant effect within the model (Table 4). Therefore, while the addition of math as a predictor led to increased fit for our model predicting movement (meaning this model had a larger adjusted R² compared to the model including only general anxiety), math anxiety itself was not a significant variable within this model. Conversely, general anxiety was a significant predictor in both models, as were math excitement and age.

Model	R	R Square	Adjusted R Square	F	p-value
1	.871 ^a	.758	.516	3.130	.078
2	.867 ^b	.752	.565	4.034	.037
3	.864 ^c	.747	.606	5.302	.015
4	.862^d	.743	.640	7.212	.005

- a. Predictors: (Constant), Gen Anx, IQ, Age, Sex, Math Excitement, Math Anx, Math Facts Fluency
- b. Predictors: (Constant), Gen Anx, Age, Sex, Math Excitement, Math Anx, Math Facts Fluency
- c. Predictors: (Constant), Gen Anx, Age, Sex, Math Excitement, Math Anx
- d. Predictors: (Constant), Gen Anx, Age, Math Excitement, Math Anx

Table 3. Modeling movement using both general and math anxiety.

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	p-value
General Anxiety ^a	(Constant)	-121.971	738.288		-.165	.872
	General Anxiety	80.355	22.124	.736	3.632	.005
	MathExcitement	62.173	27.759	.493	2.240	.049
	Age	-99.983	42.139	-.466	-2.373	.039
	IQ	6.199	5.405	.226	1.147	.278
General & Math Anxiety ^a	(Constant)	589.006	335.363		1.756	.110
	Age	-128.334	37.461	-.598	-3.426	.006
	Math Anxiety	61.100	28.838	.384	2.119	.060
	Math Excitement	77.507	22.797	.614	3.400	.007
	General Anxiety	69.001	20.599	.632	3.350	.007

a. Dependent Variable: MEA

Table 4. Coefficients for best-fitting significant models.

DISCUSSION

Our study explored the relationship between movement and anxiety in a cohort of typically developing elementary aged children. Using linear regression models with math anxiety, general anxiety, or both as predictors, we were able to determine how each anxiety type differentially relates to movement. Our results show that general anxiety had significantly positive effects on child movement during a math flashcards task, while our models including math anxiety as a predictor did not reveal any significant relationships between math anxiety and median child movement. However, the model including both forms of anxiety had the best predictive power of movement during the task, indicating both anxiety forms may distinctly contribute to movement. In all significant models, the magnitude of effect of general anxiety on movement was significantly positive, meaning that higher general anxiety scores were related to increased movement during the task. Interestingly, math excitement also had significant positive effects, meaning children who reported being more excited by math also moved more during the task. The effect of age on movement was significantly negative in our models, suggesting older children move less throughout the task. The effect of math anxiety on movement was insignificant in all models.

Our findings provide additional insight into the relationship between anxiety and movement in a group of typically developing elementary-aged children, which is an under-researched age group within anxiety research. Our results indicate a role for general anxiety, but not math-specific anxiety, in movement during a math flash cards task. While movement is a well-known manifestation of anxiety in individuals with ASD,¹⁴ these findings provide evidence that movement may also be an expression of anxiety in typically developing children. Moreover, the elicited movements may also be a means of managing feelings of anxiety, as ritualized movement has been shown to be beneficial in managing anxiety in ASD populations.¹⁴⁻¹⁶ While ritualized movements have yet to be studied as a mitigator of anxiety in typically developing cohorts, other forms of movement are known to be beneficial for reducing anxiety. In elementary students, self-reported anxiety levels decrease following moderate exercise.²² In fact, multiple studies across age groups report that in some cases, movement and exercise may be equally as beneficial as medications in managing anxiety symptoms.²³ As a result, the relationship between general anxiety and movement within our cohort may reflect that movement could also be a mechanism of managing anxiety symptoms during an anxiety-inducing activity. However, further work is needed to better understand how movement relates to anxiety and determine whether movement is simply an expression of anxiety, versus a potential soothing response mechanism in typically developing cohorts.

Our findings also highlight the exciting applications for motion-analysis software for further understanding the relationships between movement and mental states. The use of novel technology to analyze movements will allow for further subdivision and understanding of movement and its relationship to anxiety. MEA has shown promise in using movement analysis to identify individuals at higher risk for psychiatric disease,²⁴ or understanding how social interactions are physically altered among individuals with ASD.²⁵ Movement-analyzing technology also allows researchers to better understand more detailed aspects of movement, such as the rigidity or repetitiveness of specific actions, further expanding the aspects of motion that researchers are able to explore.¹⁵ As this technology is easily adaptable to various research methodologies, the application of MEA or similar technology will allow better informed exploration into the relationships between movement and anxiety in diverse populations.

While our work provides substantial insight into the relationships between anxiety forms and movement, there are multiple limitations that should be addressed in future work expanding upon this topic. The current study focused on a small cohort of predominately Caucasian typically developing children within the Charlottesville-Albemarle County Area. Future work should aim to include a larger sample with greater racial diversity. Further research should expand upon the relationship between general and math anxiety in a larger cohort to further understand the unique contributions of general and subject-specific anxiety to movement. This additional exploration should include potential age-dependent changes in the relationship between math anxiety and movement. Additionally, understanding whether increased movement in response to anxiety is compensatory, leading to better behavioral outcomes, will create applications for the classroom and other academic environments.

CONCLUSIONS

Overall, our study demonstrates how various forms of anxiety differentially manifest in elementary aged children in relation to movement. Our results show that general anxiety, but not math-specific anxiety, has important impacts on movement during a math task. These findings suggest that movement may be an indicator of anxiety in elementary aged students and raise the question as to whether movement may be a potential coping mechanism in response to general anxiety. Moreover, our findings highlight the exciting possibilities provided by Motion Energy Analysis in the examination of the relationships between movement and anxiety within students of all ages.

REFERENCES

1. Centers for Disease Control and Prevention, FastStats – Mental Health, <https://www.cdc.gov/nchs/fastats/mental-health.htm> (accessed Aug 2023)
2. Centers for Disease Control and Prevention, Anxiety and depression in children: Get the facts, <https://www.cdc.gov/childrensmentalhealth/features/anxiety-depression-children.html#:~:text=4.4%25%20of%20children%20aged%203-17%20years%20%28approximately%202.7,%28approximately%202.7%20million%29%20have%20diagnosed%20depression%20in%202016-2019> (accessed Aug 2023)
3. Goodwin, R. D., Weinberger, A. H., Kim, J. H., Wu, M., and Galea, S. (2020) Trends in Anxiety among adults in the United States, 2008–2018: Rapid increases among young adults, *J Psychiatr Res* 130, 441–446. <https://doi.org/10.1016/j.jpsychires.2020.08.014>
4. Luttenberger, S., Wimmer, S., and Paechter, M. (2018) Spotlight on math anxiety, *Psychol Res Behav Manag* 11, 311–322. <https://doi.org/10.2147/PRBM.S141421>
5. Kaswadi, K. A. B., Malik, A., and Keat, O. B. (2018) Anxiety and Academic Achievement among Undergraduate Students, *Journal of Management & Science* 16(2), 11–18.
6. Sahin, S. and Tuna, R. (2022) The effect of anxiety on thriving levels of university students during the COVID-19 pandemic, *Collegian* 29(3), 263–270. <https://doi.org/10.1016/j.colegn.2021.10.004>
7. D'Agostino, A., Schirripa Spagnolo, F., and Salvati, N. (2022) Studying the relationship between anxiety and school achievement: evidence from PISA data. *Stat Methods Appl* 31(1), 1–20. <https://doi.org/10.1007/s10260-021-00563-9>
8. Mazzone, L., Ducci, F., Scoto, M. C., Passaniti, E., D'Arrigo, V. G., and Vitiello, B. (2007) The role of anxiety symptoms in school performance in a community sample of children and adolescents, *BMC Public Health* 7:347. <https://doi.org/10.1186/1471-2458-7-347>
9. Hart, S. A. and Ganley, C. M. (2019) The Nature of Math Anxiety in Adults: Prevalence and Correlates, *J Numer Cogn* 5(2), 122–139. <https://doi.org/10.5964/jnc.v5i2.195>
10. Núñez-Peña, M. I., Suárez-Pellicioni, M., and Bono, R. (2013) Effects of math anxiety on student success in higher education, *International Journal of Educational Research* 58, 36–43. <https://doi.org/10.1016/j.ijer.2012.12.004>
11. Daker, R. J., Gattas, S. U., Sokolowski, H. M., Green, A. E., and Lyons, I. M. (2021) First-year students' math anxiety predicts STEM avoidance and underperformance throughout university, independently of math ability. *npj Sci. Learn* 6(1), 1–13. <https://doi.org/10.1038/s41539-021-00095-7>
12. Commodari, E. and La Rosa, V. L. (2021) General academic anxiety and math anxiety in primary school. The impact of math anxiety on calculation skills, *Acta Psychologica* 220:103413. <https://doi.org/10.1016/j.actpsy.2021.103413>
13. Russell, K. M., Frost, K. M., & Ingersoll, B. (2019). The relationship between subtypes of repetitive behaviors and anxiety in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 62, 48–54. <https://doi.org/10.1016/j.rasd.2019.03.006>
14. Cleveland Clinic, What is Stimming? What You Need to Know, <https://health.clevelandclinic.org/what-is-stimming/> (accessed Aug 2023)
15. Lang, M., Krátký, J., Shaver, J. H., Jerotijević, D., and Xygalatas, D. (2015) Effects of Anxiety on Spontaneous Ritualized Behavior, *Current Biology* 25(14), 1892–1897. <https://doi.org/10.1016/j.cub.2015.05.049>
16. Williams, K. L., Campi, E., and Baranek, G. T. (2021) Associations among sensory hyperresponsiveness, restricted and repetitive behaviors, and anxiety in autism: An integrated systematic review, *Research in Autism Spectrum Disorders* 83:101763. <https://doi.org/10.1016/j.rasd.2021.101763>

17. Bradley, R. S., Onovbiona, H. U., Rosario, E., and Quetsch, L. B. (2023) Current Knowledge of Emotion Regulation: The Autistic Experience, in *New Insights into Emotional Intelligence* (Rodríguez, F. M. M., Ed.) 35–50, IntechOpen.
18. Charlton, R.A., Entecott, T., Belova, E., and Nwaordu, G. (2021) “It feels like holding back something you need to say”: Autistic and Non-Autistic Adults accounts of sensory experiences and stimming, *Research in Autism Spectrum Disorders* 89: 101864. <https://doi.org/10.1016/j.rasd.2021.101864>
19. Joyce, C., Honey, E., Leekam, S. R., Barrett, S. L., and Rodgers, J. (2017) Anxiety, Intolerance of Uncertainty and Restricted and Repetitive Behaviour: Insights Directly from Young People with ASD, *J Autism Dev Disord* 47(12), 3789–3802. <https://doi.org/10.1007/s10803-017-3027-2>
20. Kapp, S. K., Steward, R., Crane, L., Elliott, D., Elphick, C., Pellicano, E., and Russell, G. (2019) ‘People should be allowed to do what they like’: Autistic adults’ views and experiences of stimming, *Autism* 23(7), 1782–1792. <https://doi.org/10.1177/1362361319829628>
21. Hamel, L. M., Moulder, R., Ramseyer, F. T., Penner, L. A., Albrecht, T. L., Boker, S., and Eggly, S. (2022) Nonverbal Synchrony: An Indicator of Clinical Communication Quality in Racially-Concordant and Racially-Discordant Oncology Interactions, *Cancer Control* 29, 1–11. <https://doi.org/10.1177/10732748221113905>
22. Philippot, A., Meerschaut, A., Danneaux, L., Smal, G., Bleyenheuft, Y., & De Volder, A. G. (2019). Impact of Physical Exercise on Symptoms of Depression and Anxiety in Pre-adolescents: A Pilot Randomized Trial. *Frontiers in Psychology*, 10. <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01820>
23. Stonerock, G. L., Hoffman, B. M., Smith, P. J., & Blumenthal, J. A. (2015). Exercise as Treatment for Anxiety: Systematic Review and Analysis. *Annals of Behavioral Medicine*, 49(4), 542–556. <https://doi.org/10.1007/s12160-014-9685-9>
24. Dean, D. J., Samson, A. T., Newberry, R., & Mittal, V. A. (2018). Motion energy analysis reveals altered body movement in youth at risk for psychosis. *Neuromotor Pathology in Schizophrenia*, 200, 35–41. <https://doi.org/10.1016/j.schres.2017.05.035>
25. Georgescu, A. L., Koeroglu, S., Hamilton, A. F. de C., Vogeley, K., Falter-Wagner, C. M., & Tschacher, W. (2020). Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: A motion energy study. *Molecular Autism*, 11(1), 11. <https://doi.org/10.1186/s13229-019-0305-1>

ABOUT STUDENT AUTHOR

Gabriella Snetkov is an undergraduate student at the University of Virginia. She will graduate with a bachelor’s degree in Neuroscience in the Spring of 2025 and plans to attend medical school afterward.

PRESS SUMMARY

Anxiety is present in the everyday life of many students and can come in different forms, such as general anxiety and math anxiety. Individuals with Autism Spectrum Disorder report that movement can be soothing for anxiety, but the relationship between anxiety and movement has not yet been explored within typically developing children. In this study, we explored how both general and math-specific anxiety relate to movement during a math flashcard task in elementary school children. We found that general anxiety uniquely contributes to the amount of movement during the task. Our results have relevance to suggesting a possible predictive relationship between movement and anxiety in typically developing children and demonstrate the possibility of movement being a coping mechanism for anxiety in these children.

Chitosan-Silver Thin Film-Coated Titanium Coupons using Silane Linkers Inhibit Biofilm and Planktonic Growth

Emily C. Montgomery^{*,}, Nidhi Gupta^{a, b,}, Matthew Atwill^{a,}, Joel D. Bumgardner^{a,} & J. Amber Jennings^a

^aDepartment of Biomedical Engineering, The University of Memphis, Memphis, TN

^bCSIR—Indian Institute of Toxicology Research, Vishvigyan Bhawan, 31 Mahatma Gandhi Marg, Lucknow 226001, Uttar Pradesh, India <https://doi.org/10.33697/ajur.2023.099>

Student: cclman22@memphis.edu*

Post-Doctoral Fellow: ngupta1@memphis.edu

Graduate Student: mtatwill@memphis.edu

Mentors: jbmgrdnr@memphis.edu, jjennings@memphis.edu

ABSTRACT

Titanium is a component of many implants and orthopedic instruments, such as screws and rods; however, this and other materials may serve as a nidus for bacterial biofilm attachment. Chitosan is a biopolymer with advantages as a surface modifier, and silver ions have broad-spectrum antimicrobial properties. For this study, chitosan is bound to silver through a novel, patented process. The purpose of this research is to characterize silane-linked chitosan-silver coatings for titanium, including comparing antimicrobial efficacy. In this study, silane-linked chitosan-silver titanium coupons reduced *Staphylococcus aureus* (*S. aureus*) viability by 98% (planktonic) and 99.5% (biofilm) while supporting viability of Saos-2 osteoblast cells at levels of 75% compared with control uncoated titanium. Due to the observation of retaining osteoblast viability while reducing bacterial viability, silane-linked chitosan-silver coatings could be useful for titanium implants to reduce post-operative infection as well as support the healing process.

KEYWORDS

Titanium; *Staphylococcus aureus*; Silver; Chitosan; Silane; Osteoblast; Antimicrobial; Coating

INTRODUCTION

Titanium implants and instruments including bone plates, hip femoral stem components, shoulder arthroplasty stems, cranial plates, and intramedullary rods are widespread in the field of orthopedics due to the material's strength, resistance to corrosion, and bone-like mechanical properties.¹ They are also employed in a variety of spine surgery applications, including pedicle screws, rods, and interbody devices.² It is understood that irregularities on the surface of the implant promote the growth of bacteria by enhancing the attachment of bacterial species on rough and irregular surfaces.³ Bacterial colonization can damage the surface TiO₂ layer. A study reported that the colonization of bacteria demonstrates prominent damage to the surface morphology and chemistry of implant surfaces.^{4,5} Moreover, once implant-associated infections occur, they can be difficult to treat, as most implants have no inherent anti-bacterial activity. Contamination of the titanium implant leads to failure in function and causes a financial burden to the patient and the health care system due to requiring revision surgery or extended antimicrobial therapy. These conditions increase the possibility for surgical complications and severely worsen patients' regular activities.^{6,7} Therefore, surface cleanliness seems to be essential for implant tissue integration. Producing coated medical implants that prevent antibiotic-resistant biofilm formation as well as inhibit planktonic viability offers health professionals and device manufacturers the potential to reduce the number and severity of postoperative infections.

Silver ions have broad-spectrum antimicrobial properties against bacteria and fungi and thus are advantageous as an implant coating.^{8,9} Silver ions affect microorganisms through different modes of action: blocking substance transport in and out of the cell, inhibiting the production of energy, promoting the generation of reactive oxygen species in the presence of oxygen, interacting with DNA to prevent replication or combine protein sulfhydryl group on bacterial membrane, and disrupt bacterial membranes to kill the bacteria.¹⁰⁻¹² Recently, some works have demonstrated that implant surfaces modified by coatings embedded with nanoparticles benefit from an enhanced healing effect and exhibit long-lasting antibacterial capability with less cytotoxic silver concentration compared to other silver structures such as microparticles and ionic compounds.^{13,14} Nano-structured surfaces usually provide multiple contact sites for the adhered bacteria, causing bacterial death via localized cell wall deformation.^{12,15} Silver nanoparticles have many advantages such as good antibacterial activity, excellent biocompatibility, and satisfactory stability against antibiotics and antibiotic organic antimicrobials for medical applications.¹⁶ An enhancement in the antimicrobial impact of silver ions is further observed in combination with biopolymer like polydopamine and calcium

phosphate.¹⁵⁻¹⁷ Particularly, chitosan is a biopolymer that is gaining widespread attention as a surface modifying agent due to its versatility, and as a combination composite chitosan has revealed great antimicrobial behavior. Additionally, chitosan seems to have the ability to induce response of T helper cells.¹⁸ Silver-loaded chitosan coating on poly (acrylic acid) coated titanium has been attempted previously via electrophoretic deposition.¹⁹ Water-soluble catechol-containing chitosan coating on titanium surfaces have been used to induce the *in-situ* reduction of Ag⁺ ions.²⁰ Recently, silver-based metal-organic framework incorporated into nanofibrous chitosan was coated on a titanium implant.²¹ Silver-loaded chitosan-heparin polyelectrolyte multilayers (PEMs) were also constructed on alkali-heat treated titanium (Ti) substrates via layer-by-layer self-assembly technique.²²

In this study, we investigated the effect of silane linking on the interaction of titanium (Ti-6Al-4V) with chitosan-silver as well as the interactions of titanium-silane-chitosan-silver composite on Saos-2 cells and *S. aureus*. The advantage of this coating is that chitosan and silver are linked through a patented process.^{23,24} This novel approach involves creating a slurry of polysaccharide in a liquid containing silver ions, filtering, washing, reducing the silver ions to atomic silver, and drying the powder.²⁴ The linking method used in this study for triethoxysilylbutyraldehyde (TESBA) to titanium is similar to a previously reported process.²⁵ It is hypothesized that chitosan-silver thin film-coated titanium coupons using silane linkers inhibit biofilm as well as planktonic growth of bacteria and simultaneously promote osteoblast proliferation.

METHODS AND PROCEDURES

Treated Coupons

Ultra-corrosion-resistant grade two titanium rod with 1" diameter was obtained from McMaster Carr Supply Co. and cut into coupons. Titanium coupons were polished with 400, 600, 800, and 1200 grit sandpaper before being sonicated in soapy water, acetone, and ethanol to remove oil and residue for 10 minutes each. The coupons were then soaked in 5M NaOH for 24h at 60°C to allow accumulation of hydroxide reactive groups on the titanium surface and rinsed with deionized (DI) water twice. The coupons were treated with triethoxysilylbutyraldehyde (TESBA) to create a 2% (v/v) silane solution in ethanol, non-adhered silane was removed with ethanol, and coupons were dried for 10 minutes in a 110 °C oven. 1% chitosan-silver solution (Chitozan Health, LLC) was added and left to dry overnight. The coated coupons were immersed in phosphate buffer for 1h, rinsed with DI water, and dried fully. Fourier-Transform Infrared Spectroscopy (FTIR) (Frontier, Perkin-Elmer, Waltham, MA, USA) was used for surface functional group characterization of the different coated coupons.

Untreated Coupons

Titanium coupons were polished with 400, 600, 800, and 1200 grit sandpaper before being sonicated in soapy water, acetone, and ethanol to remove oil and residue for 10 minutes each. The uncoated coupons were rinsed with DI water and dried fully.

Morphology

Contact angles for Ti, Ti-chitosan-silver, and Ti-silane-chitosan-silver samples were collected to understand the hydrophobicity of the coupon surface. The morphology of the samples was visualized through Scanning electron microscope (SEM) (Nova NANOSEM 650 FEI™, Hillsboro, OR, USA) after 10 nm Au-Pt coating (EMS Quorum Q150T ES plus). Energy dispersive spectroscopy (EDS) (Oxford Plus) was used to evaluate the success of the coating process by determining the presence of silver even after 24-hour aqueous exposure for washing step. Dynamic Light Scattering (DLS) (Malvern Zetasizer) was used to determine the hydrodynamic diameter of the nanoparticles.

BacTiter-Glo™ Microbial Cell Viability Assay

Bacterial viability was also assessed over time when exposed to coated coupons. Coupons were treated with chitosan-silver, and then coupons were soaked in 1X phosphate buffered saline (PBS) for either seven or fourteen days. *Staphylococcus aureus* (ATCC 29213) was statically grown in tryptic soy broth (TSB) medium overnight at 37 °C with 5% CO₂. Thereafter, coupons were inoculated in sterile well plates with approximately 10⁵ CFU of *S. aureus*. After 24 hours, coupons were removed from bacterial solution and washed with PBS to separate planktonic and biofilm samples (**Figure 1**), followed by analysis with BacTiter-Glo™ Microbial Cell Viability Assay (Promega) and luminescence readings on a Biotek plate reader with Gen5 software.

Cytocompatibility

Coupons were UV-sterilized for 20 minutes and washed in cell medium. Saos-2 cells, a cell line derived from primary osteosarcoma, were seeded at 90,000 cells/well in a sterile 12-well plate before incubation at 37 °C with 5% CO₂ and exposure to three test groups: treated coupons, untreated coupons, and tissue culture plastic (TCP) control. After 24 hours, cell viability was determined using CellTiter-Glo® Luminescent Cell Viability Assay (Promega) (n=3), and cell morphology was determined using LIVE/DEAD® Viability/Cytotoxicity Kit (Thermo Fisher Scientific Inc).

Statistical analysis

Statistical analysis was performed using GraphPad Prism 9.0.0 software (GraphPad Software Incorporation, La Jolla, CA, USA). Data was assessed using ordinary one-way analysis of variance (ANOVA) with Tukey's multiple comparisons test.

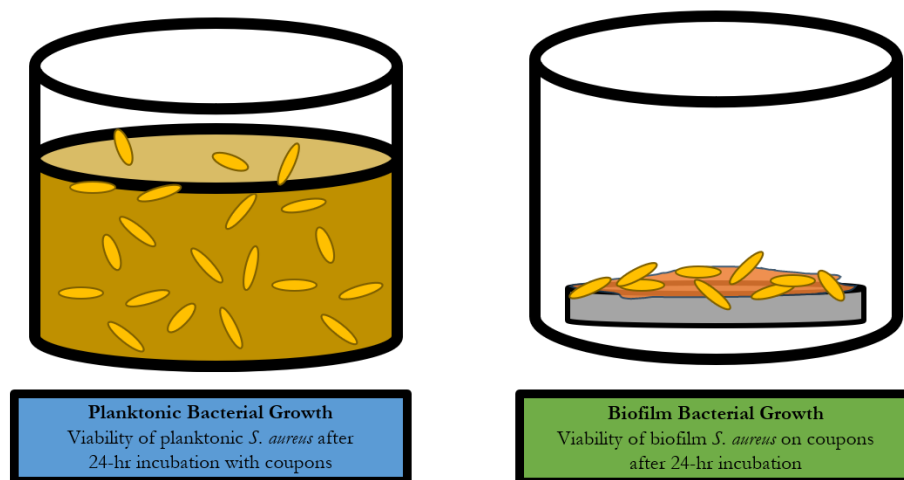


Figure 1. Division of Biofilm and Planktonic Bacteria during BacTiter-Glo Assay.

RESULTS AND DISCUSSION

Coating strategies could be implemented with many application methods including dip coating, solution casting, electrospraying, and electroplating. Of these, a simple, durable, and low-maintenance dip coating method was used in the present study. Direct titanium coating attempts with chitosan-silver solutions gave weak adherence to the substrate and delamination *i.e.* removal of the coating after contact with saline for non-silanated samples. Titanium coupons used for the study were Ti-6Al-4V, a biocompatible titanium alloy that has many medical applications and good corrosion resistance, but low antibacterial properties, which limits biological application.²⁶

Morphology

SEM/EDX analysis showed the presence of relevant elements in the coupon (**Figure S1**). Silanization of the titanium with TESBA results in uniform silane coating (**Figure S2**) that provides adherence between titanium and chitosan-silver. Chitosan-silver was well-distributed over the silanated titanium coupon surface. EDX elemental mapping for the carbon and silver fraction reflected that it corresponds to the chitosan and silver content of the chitosan-silver coating solution (**Figure 2**). Chitosan silver solution gave a Z-average diameter of 166.6 nm with polydispersity index 0.248 using DLS (**Figure S3**). A similar study using polyacrylic acid to link titanium to chitosan-silver showed homogenous coating with a surface atomic percentage of 0.3% for silver.¹⁹ Another study on the fabrication of a graphene oxide/chitosan/silver nanoparticle coating on titanium found uniform surface morphology during deposition.²⁷ Fourier transform infrared spectroscopy (FTIR) was used to examine the surface chemistry and bonding of silane and chitosan-silver to titanium coupon (**Figure 3A**). No specific peaks are observed for cleaned titanium coupon, indicating no pre-existing chemical bonds. Peaks in titanium-silane curve that appear at wavenumbers 3380–3222 cm^{-1} and 1640–1555 cm^{-1} corresponding to OH stretching and OH bending confirm the formation of Ti-OH bond on the surface.²⁸ Further, the absorbance of Si-O-Si is shown by peaks appearing at wavenumbers 1100–1000 cm^{-1} and 846 cm^{-1} confirming the success of the silanization process of the Ti surface. The FTIR spectrum of pure chitosan indicates the presence of bands at 3359 cm^{-1} (N-H and O-H stretching), 2919 cm^{-1} (C-H) 1646 cm^{-1} (NH_2 bending) and 1024 cm^{-1} (C-O-C stretching). The FTIR spectra of chitosan silver coated on silanated titanium gave a broader peak at 3337 cm^{-1} with decrease in intensity of the peak at 1646 cm^{-1} (NH_2 bending) which indicates prevalence of more O-H group as the N-H group is involved in binding to the silver metal. Generally, both O-H and N-H groups have a strong affinity towards silver ions; however, difference in electronegativity between O and N atoms plays an important role as it dictates the deprotonation site which can favor the binding of free electrons to the metal.²⁹ The presence of peak at 1643 and 1552 cm^{-1} attributed to both the OH and NH_2 bending (overlapped region due to silane and chitosan) while a shift in the Si-O-Si peaks to 1071 and 852 cm^{-1} confirms the successful coating of the chitosan silver over silanated titanium. Contact angle analysis showed an increase in angle for treated samples, potentially indicating the presence of silver (**Figure 3B-D**).

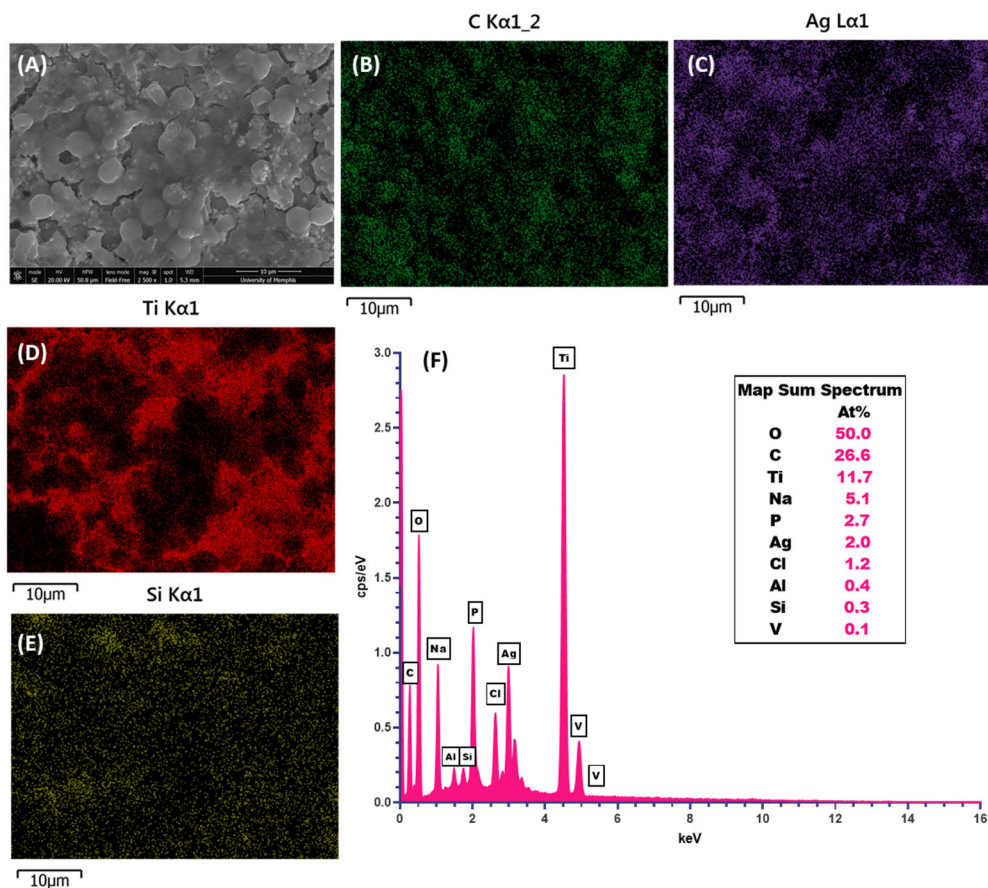


Figure 2. SEM/EDX for titanium silane chitosan silver. (A) SEM; EDX elemental mapping for (B) carbon, (c) silver, (D) titanium, (E) silica; (F) EDX spectra with atomic percentage distribution of different elements.

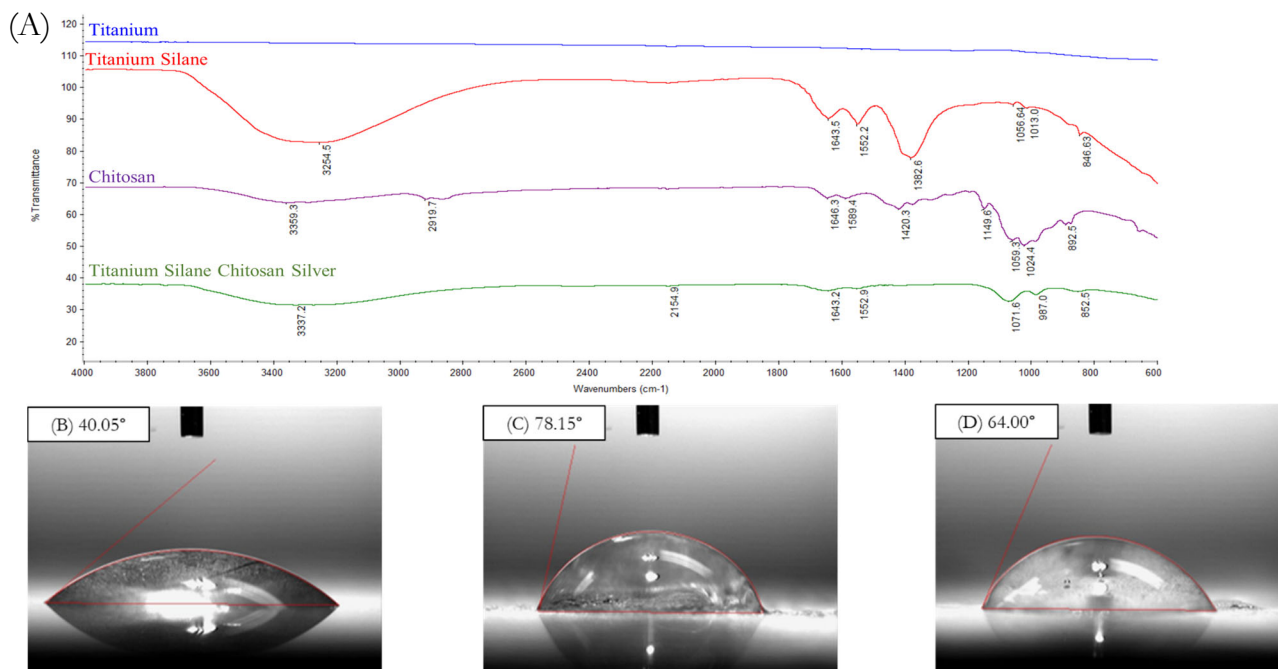


Figure 3. (A) FTIR spectra and contact angle images for (B) Ti coupon, (C) Ti-silane coupon, and (D) Ti-silane-chitosan-silver coupon.

BacTiter-Glo™ Microbial Cell Viability Assay

S. aureus is the most common bacteria responsible for prosthesis-related infections, accounting for approximately half of the infections or more.³⁰ Therefore, the antimicrobial nature of the chitosan-silver coated silanated titanium coupons was evaluated against *S. aureus*. The results of the bacteria viability assay showed that bacterial attachment was inhibited for all loaded groups (Figure 4). The reduction in planktonic viability indicates that silver is being released from the surface of the coupons at a concentration suitable for antimicrobial applications up to 14 days in PBS after loading. Overall, results suggest that the linking procedure for coating titanium surfaces promotes significant retention of silver on the titanium surface which leads to decreased planktonic and biofilm viability of *S. aureus*. Even for 14-day immersion samples, planktonic and biofilm viabilities were significantly reduced with 98% and 99.5% reductions, respectively, suggesting retention and continued elution of silver at this timepoint.

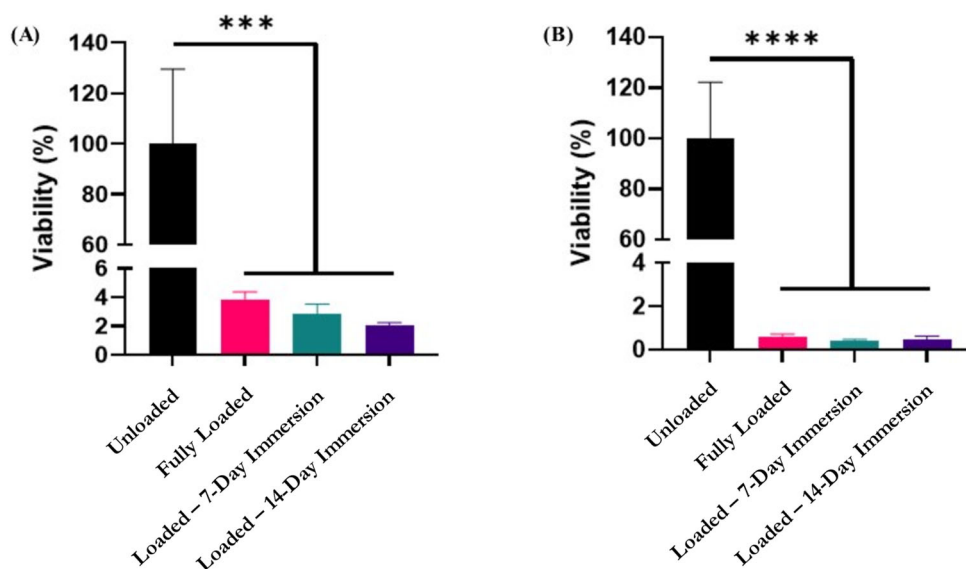


Figure 4. (A) Viability of planktonic *S. aureus* after 24-hour incubation with coupons (n=3); (B) Viability of biofilm *S. aureus* on coupons (n=3) after 24-hour incubation. Before incubating with *S. aureus*, loaded samples were immersed in PBS for 0, 7, and 14 days. **** indicates significant difference (p<0.0001) and *** indicates significant difference (p<0.001) as determined by ordinary one-way ANOVA with Tukey's multiple comparisons test.

Cytocompatibility

Silver is known to be toxic at even moderate levels, so the release of silver at non-cytotoxic levels is important in order to support cell viability.³¹ The Saos-2 cell viability for treated coupons was not statistically different than untreated coupons and was about 70 percent of the TCP control (Figure 5). Due to the sustained viability of Saos-2 cells at 70% of the control, the conjugation of chitosan with silver may also be protective against toxicity, as defined by the ISO 10993-5 Biological Evaluations of Medical Devices standard when evaluating medical devices for in vitro cytotoxicity.³² It can be observed that the addition of chitosan-silver coating does not decrease the cytocompatibility of titanium toward osteoblast cells. Silver content was greater on the surface of silanated titanium than non-silanated titanium, yet reduction in viability remained below toxic levels for both silver-containing groups.

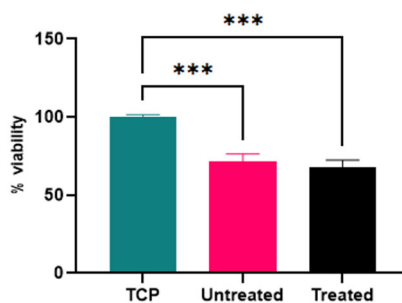


Figure 5. Percent viability of Saos-2 after 24-hour incubation with coupons (n=3). *** indicates significant difference (p<0.001) as determined by ordinary one-way ANOVA with Tukey's multiple comparisons test.

These results were also confirmed with Live/Dead staining, which produced similar results, with mostly living cells in all groups (Figure 6). Despite the presence of living cells in all groups, the treated coupons caused morphological changes in the cells and reduced the number of cells, potentially indicating apoptosis. In a similar study assessing cell response, silver nanoparticles induced significant changes in Saos-2 morphology and membrane damage.³³ Future studies could involve a lactate dehydrogenase-based or resazurin-based additional method of analysis to evaluate the long-term effects of coupons treated with chitosan-silver on Saos-2 morphology and growth as well as its *in vivo* response.

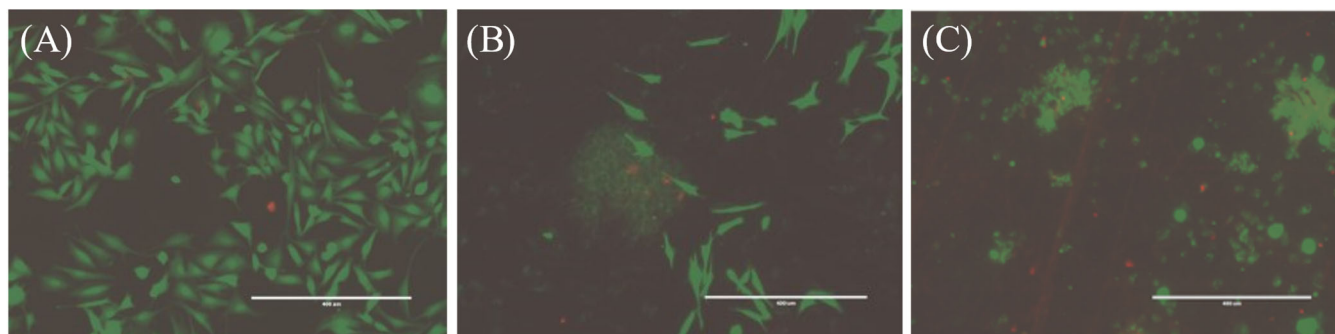


Figure 6. Live/Dead images of cells on (A) TCP control, (B) untreated coupons, and (C) treated coupons (n=1). Green and red coloration indicates live and dead cells, respectively. Scale bar is equal to 400 μm.

CONCLUSION

The dual behavior of chitosan-silver silane-linked titanium coupons to prevent biofilm and planktonic *S. aureus* growth while retaining viability of Saos-2 osteoblast cells was observed. Therefore, the approach of surface modification for titanium substrates presented here may provide an alternative strategy to simultaneously meet the desirable osteoblast growth while reducing bacterial infection for implants in clinical application.

ACKNOWLEDGEMENTS

The authors thank Ezzuddin Abuhussein for help with study completion as well as Chitozan Health, LLC for donation of chitosan-silver solution.

REFERENCES

1. Zhang, L. C., and Chen, L. Y. (2019) A Review on Biomedical Titanium Alloys: Recent Progress and Prospect. *Advanced Engineering Materials*, 21(4), 1801215. doi:10.1002/adem.201801215
2. Christensen, F. B., Dalstra, M., Sejlind, F., Overgaard, S., and Bünger, C. (2000) Titanium-alloy enhances bone-pedicle screw fixation: mechanical and histomorphometrical results of titanium-alloy versus stainless steel. *European Spine Journal*, 9(2), 97-103. doi:10.1007/s005860050218
3. Sarfraz, S., Mäntynen, P.-H., Laurila, M., Suojanen, J., Saarnio, J., Rossi, S., Horelli, J., Kaakinen, M., Leikola, J., and Reunanen, J. (2022) Effect of Surface Tooling Techniques of Medical Titanium Implants on Bacterial Biofilm Formation In Vitro. *Materials*, 15(9), 3228. doi:10.3390/ma15093228
4. Dhaliwal, J. S., David, S. R. N., Zuhlilmi, N. R., Sodhi Dhaliwal, S. K., Knights, J., and De Albuquerque Junior, R. F. (2020) Contamination of titanium dental implants: a narrative review. *SN Applied Sciences*, 2(6). doi:10.1007/s42452-020-2810-4
5. Sridhar, S., Wang, F., Wilson, T.G., Valderrama, P., Palmer, K., and Rodrigues, D.C. (2018) Multifaceted roles of environmental factors toward dental implant performance: Observations from clinical retrievals and in vitro testing. *Dental Materials* 34 (11), e265-e279. <https://doi.org/10.1016/j.dental.2018.08.299>
6. Depypere, M., Morgenstern, M., Kuehl, R., Senneville, E., Moriarty, T. F., Obremsky, W. T., Zimmerli, W., Trampuz, A., Lagrou, K., and Metssemakers, W. J. (2020) Pathogenesis and management of fracture-related infection. *Clinical Microbiology and Infection*, 26(5), 572-578. <https://doi.org/10.1016/j.cmi.2019.08.006>
7. Oikonomidis, S., Altenrath, L., Westermann, L., Bredow, J., Eysel, P., and Scheyerer, M. J. (2021) Implant-Associated Infection of Long-Segment Spinal Instrumentation: A Retrospective Analysis of 46 Consecutive Patients. *Asian Spine J*, 15(2), 234-243. doi:10.31616/asj.2019.0391
8. Soma, T., Iwasaki, R., Sato, Y., Kobayashi, T., Ito, E., Matsumoto, T., Kimura, A., Homma, F., Saiki, K., Takahashi, Y., Miyamoto, K., Matsumoto, M., Nakamura, M., M., Ishii, K., Asoda, S., Kawana, H., Xingyu, Z., Aizawa, M., Nakagawa, T., and Miyamoto, T. (2022) An ionic silver coating prevents implant-associated infection by anaerobic bacteria in vitro and in vivo in mice. *Scientific Reports*, 12(1). doi:10.1038/s41598-022-23322-6

9. Kuehl, R., Brunetto, P. S., Woischnig, A. K., Varisco, M., Rajacic, Z., Vosbeck, J., Terracciano, L., Fromm, K.M., and Khanna, N. (2016) Preventing Implant-Associated Infections by Silver Coating. *Antimicrob Agents Chemother*, 60(4), 2467-2475. doi:10.1128/aac.02934-15
10. Wang, X., Liu, S., Li, M., Yu, P., Chu, X., Li, L., Tan, G., Wang, Y., Chen, X., Zhang, Y., and Ning, C. (2016) The synergistic antibacterial activity and mechanism of multicomponent metal ions-containing aqueous solutions against *Staphylococcus aureus*. *Journal of Inorganic Biochemistry*, 163, 214-220. <https://doi.org/10.1016/j.jinorgbio.2016.07.019>
11. Lansdown, A. B. (2006) Silver in health care: antimicrobial effects and safety in use. *Curr Probl Dermatol*, 33, 17-34. doi:10.1159/000093928
12. Long, Y.-M., Hu, L.-G., Yan, X.-T., Zhao, X.-C., Zhou, Q.-F., Cai, Y., and Jiang, G.-B. (2017) Surface ligand controls silver ion release of nanosilver and its antibacterial activity against *Escherichia coli*. *International Journal of Nanomedicine*, Volume 12, 3193-3206. doi:10.2147/ijn.s132327
13. Ziabka, M., Matysiak, K., Walczak, K., Gajek, M., and Cholewa-Kowalska, K. (2022) Modification of TiAlV Alloys with Hybrid Layers Containing Metallic Nanoparticles Obtained by the Sol-Gel Method: Surface and Structural Properties. *International Journal of Molecular Sciences*, 23(4), 2283. doi:10.3390/ijms23042283
14. Ziabka, M., Matysiak, K., Cholewa-Kowalska, K., Kyzioł, A., Królicka, A., Sapiernyński, R., Januchta-Kurmin, M., and Bissenik, I. (2023) In Vitro and In Vivo Studies of Antibacterial Coatings on Titanium Alloy Implants for Veterinary Application. *International Journal of Molecular Sciences*, 24(9), 8114. doi:10.3390/ijms24098114
15. Zhang, Y., Dong, C., Yang, S., Chiu, T.-W., Wu, J., Xiao, K., Huang, Y., and Li, X. (2018) Enhanced silver loaded antibacterial titanium implant coating with novel hierarchical effect. *Journal of Biomaterials Applications*, 32(9), 1289-1299. doi:10.1177/0885328218755538
16. Jia, Z., Xiu, P., Li, M., Xu, X., Shi, Y., Cheng, Y., Wei, S., Zheng, Y., Xi, T., Cai, H., and Liu, Z. (2016) Bioinspired anchoring AgNPs onto micro-nanoporous TiO₂ orthopedic coatings: Trap-killing of bacteria, surface-regulated osteoblast functions and host responses. *Biomaterials*, 75, 203-222. doi:10.1016/j.biomaterials.2015.10.035
17. Mokabber, T., Cao, H. T., Norouzi, N., Van Rijn, P., and Pei, Y. T. (2020) Antimicrobial Electrodeposited Silver-Containing Calcium Phosphate Coatings. *ACS Applied Materials & Interfaces*, 12(5), 5531-5541. doi:10.1021/acscami.9b20158
18. Moran, H. B., Turley, J. L., Andersson, M., and Lavelle, E. C. (2018) Immunomodulatory properties of chitosan polymers. *Biomaterials*, 184, 1-9. <https://doi.org/10.1016/j.biomaterials.2018.08.054>
19. Cometa, S., Bonifacio, M. A., Baruzzi, F., De Candia, S., Giangregorio, M. M., Giannossa, L. C., Dicarolo, M., Mattioli-Belmonte, M., Sabbatini, L., and De Giglio, E. (2017) Silver-loaded chitosan coating as an integrated approach to face titanium implant-associated infections: analytical characterization and biological activity. *Analytical and Bioanalytical Chemistry*, 409(30), 7211-7221. doi:10.1007/s00216-017-0685-x
20. Cheng, Y. F., Zhang, J. Y., Wang, Y. B., Li, C. M., Lu, Z. S., Hu, X. F., and Xu, L. Q. (2019) Deposition of catechol-functionalized chitosan and silver nanoparticles on biomedical titanium surfaces for antibacterial application. *Materials Science and Engineering: C*, 98, 649-656. <https://doi.org/10.1016/j.msec.2019.01.019>
21. Dehnavi, S. M., Barjasteh, M., Ahmadi Seyedkhani, S., Rahnamaee, S. Y., and Bagheri, R. (2023) A Novel Silver-Based Metal-Organic Framework Incorporated into Nanofibrous Chitosan Coatings for Bone Tissue Implants. *Int J Pharm*, 640, 123047. doi:10.1016/j.ijpharm.2023.123047
22. Li, W., Yang, Y., Zhang, H., Xu, Z., Zhao, L., Wang, J., Qiu, Y., and Liu, B. (2019) Improvements on biological and antimicrobial properties of titanium modified by AgNPs-loaded chitosan-heparin polyelectrolyte multilayers. *Journal of Materials Science: Materials in Medicine*, 30(5). doi:10.1007/s10856-019-6250-x
23. Taylor, A., Roberts, G. A. F., and Wood, F. A. (2008) Preparation of articles having a contact biocidal property. In: Google Patents.
24. Taylor, A., Roberts, G. A. F., and Wood, F. A. (2010) Powders having contact biocidal properties. In: Google Patents.
25. Martin, H. J., Schulz, K. H., Bumgardner, J. D., and Walters, K. B. (2008) An XPS study on the attachment of triethoxysilylbutyraldehyde to two titanium surfaces as a way to bond chitosan. *Applied Surface Science*, 254(15), 4599-4605. <https://doi.org/10.1016/j.apsusc.2008.01.066>
26. Jawed, S. F., Rabadia, C. D., Khan, M. A., and Khan, S. J. (2022) Effect of alloying elements on the compressive mechanical properties of biomedical titanium alloys: A systematic review. *ACS omega*, 7(34), 29526-29542. <https://doi.org/10.1021/acsomega.2c02096>
27. Madhu, K., Kannan, S., Perumal, A., and Shanmugam, P. (2023) Biofunctionalized nanocomposite coating on Cp-titanium with reduce implant failures. *Vacuum*, 215, 112328. <https://doi.org/10.1016/j.vacuum.2023.112328>
28. Chng, E. J., Watson, A. B., Suresh, V., Fujiwara, T., Bumgardner, J. D., and Gopalakrishnan, R. (2019) Adhesion of electrospayed chitosan coatings using silane surface chemistry. *Thin Solid Films*, 692, 137454. <https://doi.org/10.1016/j.tsf.2019.137454>
29. Nate, Z., Moloto, M. J., Mubiayi, P. K., and Sibiya, P. N. (2018) Green synthesis of chitosan capped silver nanoparticles and their antimicrobial activity. *MRS Advances*, 3, 2505-2517.

30. Song, Z., Borgwardt, L., Høiby, N., Wu, H., Sørensen, T. S., and Borgwardt, A. (2013) Prosthesis infections after orthopedic joint replacement: the possible role of bacterial biofilms. *Orthopedic Review*, 5(2), 14. doi:10.4081/or.2013.e14
31. Tortella, G. R., Rubilar, O., Durán, N., Diez, M. C., Martínez, M., Parada, J., and Seabra, A. B. (2020) Silver nanoparticles: Toxicity in model organisms as an overview of its hazard for human health and the environment. *Journal of hazardous materials*, 390, 121974. <https://doi.org/10.1016/j.jhazmat.2019.121974>
32. I. Standardization (1999) ISO-10993-5: Biological Evaluation of Medical Devices Part 5: Test for Cytotoxicity: In Vitro Methods, ANSI/AAMI: Arlington, VA, USA.
33. Sokolowska, P., Siatkowska, M., Bialkowska, K., Rosowski, M., Komorowski, P., and Walkowiak, B. (2022) Osteosarcoma cells in early and late stages as cancer in vitro progression model for assessing the responsiveness of cells to silver nanoparticles. *Journal of Biomedical Materials Research Part B: Applied Biomaterials*, 110(6), 1319-1334. doi:10.1002/jbm.b.35002

ABOUT THE STUDENT AUTHORS

This work began when E.C. Montgomery was a senior undergraduate student. E.C. Montgomery graduated with a bachelor's degree in May 2022. Since that time, E.C. Montgomery has completed a master's degree in biomedical engineering and continued work under the supervision of the mentors mentioned.

PRESS SUMMARY

Titanium implants and instruments are important and commonly used in medicine because of the material's strength; however, titanium can encourage bacteria to attach to an implant surface, forming a biofilm, which complicates infection treatment. A common bacteria involved in these implant infections is *Staphylococcus aureus*. Chitosan is a natural material found in crustacean shells, and silver ions have broad-spectrum antimicrobial properties. The purpose of this research is to test if titanium coated with chitosan-silver reduces biofilm attachment compared to titanium alone. Specifically, in this study, chitosan-silver titanium coupons reduced *S. aureus* viability by 99%, compared to uncoated titanium, without significantly affecting the growth of bone cells. Due to this finding, silane-linked chitosan-silver coatings on titanium implants could reduce implant infection and support natural bone healing.

