AJUR

American Journal of
Undergraduate Research

Volume 20 | Issue 2 | September 2023

www.ajuronline.org

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

AJUR
American Journal of
Undergraduate Research

Volume 20 | Issue 2 | September 2023 | https://doi.org/10.33697/ajur.2023.082

2 **AJUR History and Editorial Board**

3 **An Exploratory Study on Student-Athlete Mental Health: Personal and Perceived Barriers to Help-Seeking Behavior**
Emma McCabe, Sarah DeSordi, Aaron Piepmeier, & Eric Hall

13 **The Effect of Coastline Concavity on Maximum Storm Surge Height along the US Gulf Coast**
Kayleigh Addington & Stephanie Zick

29 **Validation of a Computationally Efficient Model of the Mu-Opioid Receptor**
Allison Barkdull, Lexin Chen, Akash Mathavan, Karina Martinez-Mayorga, & Coray M. Colina

45 **A Review of Models on Direct Evaporative Cooling**
Michael Wilkins & Nelson Fumo

57 **Semantic Interpretations of Ditransitive Constructions in English**
Marcella Jurotich

69 **Are Wrist-based Heart Rate Monitors a Valid Tool for Fitness Professionals to Measure Training Intensity During Exercise Classes?**
Korey Little, John C. Sieverdes, D. David Thomas, M. Blake Lineberger, Daniel B. Bornsteind, Marco Bergamine, & Wesley D. Dudgeon

79 **Retributive Attitudes and Perceptions of Police Use of Excessive Force**
Amelia Collins, Sherah L. Basham, & Rick Dierenfeldt

87 **College Canines: Investigating the Behavioral and Physiological Impacts of Various College-Housing Environments on Companion Dogs**
Kaitlyn Willgohs, Jenna Williams, Isabella Crisostomo, Katherine Keck, Crystal Young-Erdos, & Lauren Highfill

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 cheritable not-for-profit organization in 2018. AJUR is indexed internationally by EBSCO and Crossref with ISSNs of 1536-4585 (print) and 2375-8732 (web).*

An Exploratory Study on Student-Athlete Mental Health: Personal and Perceived Barriers to Help-Seeking Behavior

Emma McCabe[a,b], Sarah DeSordi[a], Aaron Piepmeier[a], & Eric Hall[a]*

[a]Department of Exercise Science, Elon University College of Arts and Sciences, Elon, NC
[b]National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD

https://doi.org/10.33697/ajur.2023.083

*Students: emma.mccabe@mail.nih.gov, sdesordi@elon.edu**
Mentors: apiepmeier@elon.edu, ehall@elon.edu

ABSTRACT
Student-athletes are more likely to develop mental health problems than the general population. In addition to schoolwork, social networks, family ties, and financial obligations, collegiate student-athletes are required to attend practices, travel for games, attend athletic events, and perform extracurricular duties. The addition of possible injury, overtraining, burnout, scrutiny from the public and/or the media, and consistent pressure to perform results in athletes playing through both physical injuries and mental health problems. Despite the high number of athletes who report needing mental health support, fewer than half seek out mental health services. Research has identified stigma as one of the largest barriers to mental health help-seeking behavior (HSB). Help-seeking behavior has rarely been studied in relation to the larger body of work on mental health stigma in sport. The purpose of this study was to observe and describe student-athletes' perceived stigma (e.g., what others think) and personal stigma (e.g., what the individual thinks) in relation to HSB. A sample of $n = 20$ athletes completed an online Qualtrics Survey, which included Link's Perceived Discrimination and Devaluation Scale, Mental Health Literacy Scale, Self-Stigma of Seeking Help Scale, Help Seeking Questionnaire, and Student-Athlete Role Behaviors Questionnaire. Results from this study may help develop mental health interventions to improve student-athlete HSB.

KEYWORDS
Mental Health; Student-Athlete; Stigma; Help-Seeking Behavior; NCAA; PDDS; MHLS; SSOSH; HSQ; SRBQ

INTRODUCTION

Mental health disorders account for an estimated $800 billion in direct costs and $1.7 trillion in productivity loss globally per year.[1] Over 50% of the world's population will experience a mental health disorder at some point in their lives.[1] From 1990 to 2017, the global incidence of depression rose more than 49%.[2] The COVID-19 pandemic has only exacerbated mental health problems worldwide.[3—6] A population study from 2007 to 2017 found that US college students had a 15% increase in utilization of mental health resources.[7] Lifetime diagnoses of depression and suicidal ideation also saw a 14% increase.[7] Increased utilization of resources and decreases in stigma are noted as potential reasons for upticks in diagnoses among the general college student population.[7]

Student-athletes are at greater risk for mental health problems compared to the general population due to a unique series of factors resulting from their athletic status.[8—12] Student-athletes are under constant pressure to perform athletically, and a demand for "mental toughness" above all else often results in athletes playing through both physical injuries and mental health problems.[13—16] Research shows that while over 54% of Division I (DI) National Collegiate Athletic Association (NCAA) student-athletes have stated a need for mental health support, fewer than half have utilized mental health resources.[9]

The NCAA lists student-athlete mental health as a main priority in its mission statement.[17] In addition to the schoolwork, social networks, family ties, and financial obligations juggled by the average college student, collegiate student-athletes are required to attend practices, travel for games, attend athletic events and perform extracurricular duties.[8, 11, 13] The addition of possible injury, overtraining, burnout, scrutiny from the public and/or the media, and consistent pressure to perform contributes to mental health problems.[8, 12—13] NCAA recruiting starts before college, meaning young athletes enter into competitive environments often before developing proper coping mechanisms.[13] Poor mental health outcomes can result in a stressful environment without assistance from mental health resources.[13] Chronic stress can have detrimental effects on brain structure and function, including reductions in hippocampal volume, which are also notable in mental health disorders like anxiety and depression.[18—19]

Prejudice and discrimination may prevent those labeled with mental health problems from achieving individual-level success and independence, regardless of athletic status.[15] Stereotypes that those with mental health problems are inherently dangerous, incapable, or immoral can lead to discriminatory behaviors, like missed job or educational opportunities and social exclusion.[15] For example, a societal stereotype of individuals with depression might be that they are lazy, have few goals, and are incapable of achieving those goals. This negative judgment of others that can result in social ostracism is called stigma.[15] Prejudice would be believing in these stereotypes (e.g., *I agree, people with depression are lazy and I don't want to play on a team with them. They will put forth little effort and we will lose*).[15] Such biases can result in discrimination, or actions that prevent social advancement by the afflicted group (e.g., *This soccer player has depression, so I will not let them play in games or involve them in team activities*).[15] If those in the outcast group learn to believe these stereotypes, they may begin to hold negative views toward themselves that lend to poor self-esteem, self-sabotaging behaviors, or a learned helplessness that can prevent social advancement (e.g., *Society sees people with depression as lazy. I agree, people with depression are lazy. I have depression. Therefore, I am lazy and cannot achieve my goals. No one else thinks I can either, so why try?*).[15] Negative judgment held toward oneself is termed self-stigma.[20—22]

"Help-seeking behavior" has become a popular term to describe intentional, planned behavior that involves seeking care from a healthcare service or professional.[23] Utilization of professional and interpersonal connections or resources to address one's mental health (e.g., counseling, prescribed medication) is defined as mental health help-seeking behavior (HSB).[15, 23—24] Stigma is regarded as a well-known barrier to seeking mental health services.[15] Student-athletes may avoid HSB for fear of coaches or teammates viewing them as incapable, which could result in lack of playing time and/or social rejection.[9, 25—27] Even if these fears can be proven false, student-athletes' consciousness of societal bias towards those with mental health problems can prevent HSB. The beliefs that an individual holds towards those with mental health problems (e.g., *People with depression are lazy*) is defined as personal stigma.[15, 21, 24] The interpretation of others' personal stigmas is defined as perceived stigma (e.g., *Other people think people with depression are lazy*).[15, 21, 24] The internalization of personal and perceived stigmas can result in self-stigma.

It is largely unknown how student-athletes' perceived stigma compares with the personal stigma of their peers or their own personal stigma in regard to HSB.[24] Results from a previous study comparing perceived and personal stigma found that student-athletes reported greater perceived stigma than personal stigma, and that student-athletes reported greater overall stigma than their non-athlete peers.[24] Student-athletes from a similar study reported less perceived stigma than non-athletes, though neither group differed in terms of HSB.[28] Athletes often cite perceived stigma and low mental health literacy as barriers to HSB.[9, 25—27] Past studies suggest that perceived benefits of seeking help and self-efficacy, both of which are directly influenced by perceived stigma and low mental health literacy, are important factors of help-seeking.[28—30] Help-seeking behavior has rarely been studied in relation to the larger body of work on mental health stigma in sport.[9, 20, 24—27] While Hillard et al. found a negative relationship between student-athletes with greater mental health stigma and likelihood of seeking professional help and another study demonstrated that athletes' self-stigma was negatively related to HSB, none of the literature found discussed or surveyed relative barriers to HSB in congruence with perceived, personal, and self-levels of stigma in this population.[31—32]

The purpose of this study was to observe and describe student-athletes' personal and perceived stigma in relation to HSB (Study 1). Participants completed an online survey that assessed their perceived stigma, personal stigma, self-stigma, present and past HSB, and athlete behaviors, with the goal of comparing perceived and personal stigma. A similarly conducted study from 2019 was analyzed (Study 2). Results from these studies may inform future hypotheses as well as methods to improve HSB among student-athletes. We hypothesize that student-athletes' perceived stigma will be greater than their personal stigma.

Study 1
METHODS AND PROCEDURES

Participants

Twenty student-athletes (80% female) from a southeastern private university participated in this study between November and April of the 2020-21 academic year. All study participants were White, and sexual orientation was reported as follows: heterosexual (90%), bisexual (10%). The teams most represented were soccer (20%) and either Cross-Country and/or Track-and-Field athletes (25%) (see Supplementary Materials for demographic breakdown by sport). Inclusion criteria were: full-time student status (at least a 12-semester hour course load), between the ages of 18 and 24 years, and self-identified as having participated in NCAA varsity athletics during the past 12 months.

Measures

The Link's Perceived Discrimination and Devaluation Scale (PDDS) is a 12-item survey that assesses one's perceived stigma of individuals who have sought psychiatric help for mental health problems.[21] Responses are based on a 5-point Likert-scale, ranging from (1) *Totally Agree* to (5) *Totally Disagree,* and scores are added to a total score out of 60. A higher score on the PDDS reflects a greater perception that those with mental health problems will be discriminated against and devalued by the student-athlete peer

group. The PDDS is valid and reliable for student-athlete populations, and previous studies with student-athletes have modified the wording of the subject of the questions to fit the specific peer-group population.[24] Therefore, we employed these modifications. Participants ranked their level of agreement with statements such as, "Most *of my fellow student-athletes* would willingly accept someone who has received mental health treatment as a close friend."

The Mental Health Literacy Scale (MHLS) is a 35-item questionnaire that assesses participants' knowledge, understanding, and stigma attributed to mental health.[33] The total score for the MHLS was determined by summing the scores to a total score out of a possible 175, with a minimum score of 35.

The Self-Stigma of Seeking Help Scale (SSOSH) is a 10-item instrument that assesses participants' self-stigma and self-esteem related to HSB.[22] Items are measured on a 5-point Likert scale with (1) meaning *Strongly Disagree* and (5) meaning *Strongly Agree.* The scale has been found to be reliable and valid.[22] Average responses to each item were averaged among all participants, where a higher score out of 5 indicated greater self-stigma.

A modified version of the 18-item Help-Seeking Questionnaire (HSQ) was used to assess participants' likelihood to seek help for a mental health problem, as well as any past help-seeking behaviors.[34] The first five questions assess participant's likelihood to seek help from a mental health professional, using a 7-point Likert-scale, ranging from (1) *Worthless* to (7) *Valuable,* (1) *Unpleasant* to (7) *Pleasant* (Item 1 only), (1) *Should Not* to (7) *Should* (Item 2), (1) *Disapprove* or (7) *Approve* (Item 3), or (1) *No Control* to (7) *Total Control* (Item 4). Questions 6 and 7 ask if and when the participant has ever received a mental health diagnosis. The remaining 11 questions require participants to reflect on any past experiences with mental health problems and related help-seeking behaviors they have exhibited. If they have received help in one area, participants are then asked to reveal who prompted them to seek out that resource (themselves, a loved one, or someone else), and to rank how helpful the resource was for their mental health.

The HSQ was scored by summing the responses to the first 5 questions to a total out of 35. The diagnosis questions were reported in the demographics table (see Supplementary Materials). For the third section of the HSQ, participant responses were assigned numerical values based on whether they had (1) or had not (0) used a particular resource. The reported average use for each mental health resource was the result of the proportion of those who stated that they had ever used the resource, out of the total number of participants who answered that question. Those proportions were averaged to get a total score for overall use of all mental health resources.

The Student-Athlete Role Behaviors Questionnaire (SRBQ) is an 8-item self-report instrument built to assess the behaviors associated with being a student-athlete (Lopez, 2008). Questions in this survey consider factors that have been shown to relate to depression, such as time commitment, destructive behaviors, performance potential, and roles. Items were scaled using a 7-point Likert scale ranging from (1) *Never* to (7) *Always* (Items 1 and 2) or from (1) *Strongly Disagree* to (7) *Strongly Agree* (Items 3 through 8). Answers were summed out of a possible 56.

The Feasibility Questionnaire is an ad-hoc, 5-item self-report instrument that assesses participant responses to questions about: difficulty completing the survey, pleasure experienced from completing the survey, and likelihood of suggesting or passing the survey along to another person. Participants select an answer choice on a 6-point Likert scale ranging from (1) *Strongly Disagree* to (2) *Strongly Agree*. Total scores were averaged and will be considered when refining the study design for a follow-up or future experiment.

Procedures

An interest email with a description of the study and link to the online screening survey was sent to all student-athletes. Participant names and emails were gathered from the roster lists for each team listed on the publicly available University Athletics website (https://elonphoenix.com/). Those interested in the study were asked to complete a screening survey to determine eligibility. The study protocol was approved by the Institutional Review Board of Elon University (Elon, NC).

Eligible participants were sent an email with a unique ID number, as well as a link to the online Informed Consent and Official Student-Athlete Mental Health Qualtrics Survey (Qualtrics, Provo, UT and Seattle, WA). After signing consent, participants completed an approximately 25–30-minute Qualtrics Survey (Qualtrics, Provo, UT and Seattle, WA), which included a Demographic Survey, PDDS, MHLS, SSOSH, HSQ, SRBQ, and a Feasibility Survey. All survey responses were collected and stored in a Qualtrics account (Qualtrics, Provo, UT, and Seattle, WA). Participants received a link to the project poster presentation after the study was completed.

RESULTS

Table 1 shows the summed average scores for all measures. After summing the individual responses across all of the PDDS questions, Study 1 participants had a total perceived stigma rating of 38.60 ± 3.86. Average response values to individual items revealed that most participants selected mainly 4 or 5 on a 5-point Likert scale. For example, student-athletes in Study 1 averaged 4.70, indicating that participants *Totally Disagree* with the statement, "most of my fellow student-athletes would willingly accept someone who has received mental health treatment as a close friend" (see Supplementary Materials).

Measure	Study 1 M (SD)
Perceived stigma (PDDS)	38.60 ± 3.86
Personal stigma (MHLS)	107.05 ± 9.09
Self-stigma (SSOSH)	3.20 ± 0.79
Likelihood to seek help (HSQ)	12.95 ± 2.86
Overall Use of MH Resources (HSQ)	0.32 ± 0.33
Student-athlete behavior (SRBQ)	27.35 ± 7.41
Satisfaction with survey (Feasibility)	23.95 ± 4.10

Table 1. Summed averages and standard deviations for PDDS, MHLS, SSOSH, HSQ, SRBQ, and Feasibility questionnaires.

Study 1 participants' MHLS score was around 107.05 ± 9.09. There were four section themes for the MHLS, including two sections for personal stigma. One section gave a score for personal beliefs about mental health and those with mental health problems. The average score for this section was 38.75 ± 4.08 out of a potential 45. The other section gave a score for an individual's willingness to interact with those who have a mental health problem. The average score for this section was 30.70 ± 4.67 out of a potential 35.

Based on the magnitude of response values to individual items on the PDDS compared with the MHLS, Study 1 participants were determined to have greater perceived stigma than personal stigma.

The total average SSOSH score for Study 1 was 3.20 ± 0.79.

The average participant score for Study 1 participants on the first 5 items of the HSQ was 12.95 ± 2.86. About 35% of Study 1 participants reported a prior mental health diagnosis in the demographics questionnaire. The average use of all resources for Study 1 participants was 0.32 ± 0.33, with the most used resources being "Individual Therapy/Counseling" (0.91) and "Medication from your Primary Care Physician" (0.64). Of those who endorsed seeking help, the most common stakeholders were either "my loved ones" or themselves ("I did"). NCAA Division I student-athletes reflected low perceived value, pleasantness, social approval, and level of control over their prospective HSB.

The total average SRBQ score for Study 1 participants was 27.35 ± 7.41. At least half of participants reported in the SRBQ feeling as though involvement in collegiate athletics takes away from their free time, academic potential, and causes suffering in other areas of life. About three-quarters of participants (0.75) *Agree* or *Strongly Agree* to engaging in self-destructive behaviors like binge-drinking, drug use, physical or verbal altercations, self-mutilation, or cutting to cope with their emotions (Question 5). It is important to note that specific modalities of self-destructive behaviors participated in were not identified through this survey.

Around half of the participants reported that they either *Somewhat Agreed, Agreed*, or *Strongly Agreed* that they could hide having a mental health problem from their coaches and teammates (Question 8).

Study 2

METHODS AND PROCEDURES

Participants

PDDS and HSQ data from 54 student-athletes (83% female) was collected from the same private university in 2019. The majority of participants were White (83%), Black/African American (5%), and mixed race (12%) (see Supplementary Materials for detailed breakdown). Sexual orientation was: heterosexual (91%), bisexual (6%), homosexual (4%), other (2%). Approximately 28% of Study 2 participants were soccer athletes, while 19% of players reported being from "other" sports (see Supplementary Materials for demographic breakdown by sport). This unpublished dataset had a similar demographic makeup to Study 1 and could reduce the chance of bias within our results.

Measures

The only measures from Study 2 discussed in this paper were the PDDS and HSQ (see Study 1 Measures).

Procedures

A single email with a link to a Qualtrics survey was sent to all student-athletes at the university. Only the PDDS and HSQ are described in this paper. It should be noted that the PDDS used in Study 2 maintained the "Most people" phrasing of the original questionnaire.[18]

RESULTS

Table 2 shows Study 2 summed average scores for PDDS and HSQ. Study 2 participants had a total perceived stigma rating of 40.81 ± 2.76. The average participant score for Study 2 participants on the first 5 items of the HSQ was 19.52 ± 2.58. The HSQ used in Study 2 did not include mental health diagnosis questions, but demographics revealed that at least 41% participants had received a previous mental health diagnosis. Study 2 participants' average use of all resources was 0.15 ± 0.20, with the most used resources being "Individual Therapy/Counseling" (0.51) and "Consulting Family/Friends" (0.49).

Measure	Study 2 M (SD)
Perceived stigma (PDDS)	40.81 ± 2.76
Likelihood to seek help (HSQ)	19.52 ± 2.58
Overall Use of MH Resources (HSQ)	0.15 ± 0.20

Table 2. Summed averages and standard deviations for PDDS and HSQ.

DISCUSSION

The purpose of this study was to describe athletes' perceived stigma and personal stigma in relation to HSB. NCAA Division I student-athletes demonstrated higher levels of perceived stigma than personal stigma, about average mental health literacy, high self-stigma, and low HSB. No notable differences were found between those with or without reported mental health diagnoses in either dataset, amongst all questionnaires. Our results strongly reflect the current literature. It would make sense that our student-athlete population would avoid HSB if they believe that being open about mental health problems will lead to rejection from their peers, as indicated by high perceived stigma scores.

A previous study found a strong relationship between student-athlete role behavior and outward expression of depression.[35] Results noted that student-athletes were likely to reject treatment, express low perceived need to change, and low motivation to seek help.[35] It is concerning that the majority of Study 1 athletes regularly engage in "self-destructive" behaviors, including binge-drinking, drug use, physical or verbal altercations, self-mutilation, or cutting. Self-destructive behaviors are often coping mechanisms for poor mental health but may not be indicative of a diagnosable mental health problem.[3, 36] Since the survey generalizes self-destructive behaviors, it is unclear specifically what type(s) of self-harm is being done.

Sullivan et al. discovered that the average MHLS score for athletic staff resembled that of student-athletes at around 131.48, which was deemed "poor."[37] Study 1 student-athletes scored lower than this norm (107.05), and by that standard demonstrated

poor mental health literacy. Though the MHLS has no official ranking system, Study 1 student-athletes scored above the minimum 35 score on the MHLS and just greater than the mean score, suggesting about average or slightly above average mental health literacy.[33] More work needs to be done to eliminate discrepancies in scaling the MHLS, specifically in student-athlete populations. Higher mental health literacy is related to greater HSB, so Study 1 student-athletes may demonstrate lower intentions to seek help, as reflected in the HSQ results.[33] Despite lower mental health literacy, it is encouraging that Study 1 student-athletes scored over 30.70 $\pm$ 4.67 out of 35 on the MHLS sections referencing personal stigma. As supported by the SSOSH data, more often than not, NCAA Division I student-athletes hold greater self-stigma around HSB. Players that are more accepting of those with mental health problems (e.g., have low personal stigma) may be able to reduce the level of fear surrounding HSB.

Literature has found that, besides direct family members, student-athletes are more likely to partake in HSB if referred by a coach than a teammate or peer.[11, 38] As student-athletes are less likely to interact with family on a daily basis compared with coaches or teammates, coach attitudes toward mental health are extremely influential for student-athletes.[2, 11, 13, 38—40] The fact that athletes feel capable of hiding mental health problems from team members may be explained by a combination of factors like stigma, poor team cohesion or relationships, player family history and culture, and perceived coach attitudes.[11, 41—43] Though few studies have explored nondisclosure in terms of mental health, recent research on concussion symptom reporting demonstrated that greater concussion knowledge was associated with greater nondisclosure, and that student-athletes were less likely to report symptoms in high-stakes situations.[44] Similar contexts may influence the desire for athletes to disclose symptoms of mental health problems. Future hypotheses could test athlete-stakeholder relationships.

We limited our study to focus on student-athletes alone and did not include coaches, athletic trainers, and other stakeholders. Improving coach mental health literacy, awareness of student-athletes' perceptions of HSB, and communication with players may be important steps in improving athletes' HSB.[11, 13, 39—40] Given that student-athletes endorse HSB when referred by family or themselves on the HSQ and they are likely to spend most hours with coaches and teammates, next steps should involve coach-athlete interventions that aim to improve mental health literacy, reduce stigma, and generally improve HSB. Successful mental health interventions have incorporated knowledge and self-efficacy training with both athletes and training staff to boost mental health awareness.[28—30, 45] A recent study that put athletes, coaches, and team captains through mental health training with a focus on social support and coping resulted in improved athletic coping skills and reduced anxiety amongst players.[45] Studies using brief educational presentations have gotten positive feedback from athletes.[46—47] While not all interventions yielded significant increases in immediate help-seeking attitudes and behavior relative to controls, greater increases in HSB could be found post-intervention.[48] Generally, student-athletes are more likely to seek help for a mental health problem after participating in mental health literacy and HSB interventions.[8, 45, 47—48]

This was an exploratory study where exclusion criteria were kept minimal to allow for as large a sample size as possible. Because the Study 1 sample size was demographically homogeneous and small (n = 20), there was not enough power in the sample to justify statistical tests that would explore differences between variables, both in Studies 1 and 2. For this reason, we did not control for mental health diagnoses during statistical analysis. This study should be repeated with a larger, more representative sample. For example, after scoring the PDDS and HSQ from both studies, it appears that student-athletes from Study 2 demonstrate lower levels of perceived stigma (PDDS) and a higher HSB (HSQ). Without data to confirm, we theorize that these discrepancies may be attributed to any of the following factors: sample size, the time at which the participants took the assessment (e.g., pre- vs. mid-pandemic), and demographic differences.

Another limitation to combining datasets was the timing at which each survey was conducted. Study 1 was open between October of 2020 and April of 2021. This overlapped with the COVID-19 global pandemic, in which the majority of university operations were conducted virtually, and many varsity sports were interrupted or altered.[5, 49—50] Study 2 took place in 2019, before the pandemic. The COVID-19 pandemic is a global event which has placed varying degrees of stress on all individuals and has increased the mental health burden worldwide.[4, 6, 51] It is possible that the responses given by Study 1 student-athletes were influenced by their experiences while living through a global pandemic.[52—53]

CONCLUSIONS

This pilot study investigated differences in personal and perceived stigma in NCAA Division I student-athletes in relation to HSB. Student-athletes exhibited higher levels of perceived stigma than personal stigma, about average mental health literacy, high self-stigma, and low HSB. Despite reporting less personal stigma, student-athletes expressed little desire to seek help for their own mental health. Future studies could examine which type(s) of stigma contribute most significantly to HSB. Our results strongly reflect the current literature, which may inform future mental health interventions geared toward lowering stigma and increasing mental health literacy.

ACKNOWLEDGMENTS

The authors thank Kennedy Little (Elon University, Class of 2020) and James Contino (Elon University, Class of 2019) for the collection of student-athlete data for Study 2.

REFERENCES

1. Trautmann, S., Rehm, J. and Wittchen, H.U. (2016) The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? *EMBO Rep,* 17, 1245–1249. *https://doi.org/10.15252/embr.201642951*
2. Liu, Q., He, H., Yang, J., Feng, X., Zhao, F. and Lyu, J. (2020) Changes in global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study, *J Psychiatric Res*, 126, 134–140. *https://doi.org/10.1016/j.jpsychires.2019.08.002*
3. Koob, G.F., Powell, P. and White, A. (2020) Addiction as a coping response: Hyperkatifeia, deaths of despair, and COVID-19, *Am J Psychiatry*, 177(11), 1031–1037. *https://doi.org/10.1176/appi.ajp.2020.20091375*
4. Racine, N., McArthur, B.A., Cooke, J.E., Eirich, R., Zhu, J. and Madigan, S. (2021) Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: A meta-analysis, *JAMA Pediatr,* 175(11), 1142–1150. *http://jamanetwork.com/article.aspx?doi=10.1001/jamapediatrics.2021.2482*
5. Walke, H.T., Honeien, M.A. and Redfield, R.R. (2020) Preventing and responding to COVID-19 on college campuses, *JAMA. https://doi.org/10.1001/jama.2020.20027*
6. Wu, T., Jia, X., Shi, H., Niu, J., Yin, X., Xie, J. and Wang, X. (2021) Prevalence of mental health problems during the COVID-19 pandemic: A systematic review and meta-analysis, *J Affect Disord,* 281, 91–98. *https://dx.doi.org/10.1016%2Fj.jad.2020.11.117*
7. Lipson, S.K., Lattie, E.G. and Eisenberg, D. (2018) Increased rates of mental health service utilization by U/S/ college students: 10–year population-level trends (2007-2017), *Psychiatr Serv*, 70(1), 60–63. *https://doi.org/10.1176/appi.ps.201800332*
8. Barrasso, M.S. (2020) Mental health and student athletes, *Community Engagement Student Work,* 38. *https://scholarworks.merrimack.edu/soe_student_ce/38*
9. Moore, M. (2017) Stepping outside of their comfort zone: Perceptions of seeking behavioral health services amongst college athletes, *J Issues Intercoll Athl,*130–144. *http://csri-jiia.org/wp-content/uploads/2017/07/JIIA_2017_SI_08.pdf*
10. Moreland, J.J., Coxe, K.A. and Yang, J. (2018) Collegiate athletes' mental health service utilization: A systematic review of conceptualizations, operationalizations, facilitators, and barriers, *J Sport Health Sci*, 7(1), 58–69. *https://doi.org/10.1016/j.jshs.2017.04.009*
11. Reich, A.L. (2019) Memorable messages that shape student-athletes' perceptions of seeking mental health services, *ProQuest Dissertations Publishing. http://libres.uncg.edu/ir/uncg/f/Reich_uncg_0154D_12820.pdf*
12. Rice, S.M., Purcell, R., De Silva, S., Mawren, D., McGorry, P.D. and Parker, A.G. (2016) The mental health of elite athletes: A narrative systematic review, *Sports Med,* 46(9), 1333–1353. *https://doi.org/10.1007/s40279-016-0492-2*
13. Bauman, N.J. (2016) The stigma of mental health in athletes: Are mental toughness and mental health seen as contradictory in elite sport?, *Br J Sports Med,* 50(1), 135–136. *http://dx.doi.org/10.1136/bjsports-2015-095570*
14. Carr, C. and Davidson, J. (2015) Mind, body, and sport: The psychologist Perspective. NCAA. *https://www.ncaa.org/sports/2014/11/3/mind-body-and-sport-the-psychologist-perspective.aspx*
15. Corrigan, P.W., Druss, B.G. and Perlick, D.A. (2014) The impact of mental illness stigma on seeking and participating in mental health care, *Psychol Sci Public Interest,*15(2), 37–70. *https://doi.org/10.1177/1529100614531398*
16. Wiese-Bjornstal, D.M., Smith, A.M., Shaffer, S.M. and Morrey, M.A. (1998) An integrated model of response to sport injury: Psychological and sociological dynamics, *J Appl Sport Psychol*, 10(1), 46–69. *https://doi.org/10.1080/10413209808406377*
17. National Collegiate Athletic Association, Mission and Priorities, *https://www.ncaa.org/sports/2021/6/28/mission-and-priorities.aspx* (accessed Oct 2022).
18. Lupien, S.J., Juster, R.P., Raymond, C. and Marin, M.F. (2018) The effects of chronic stress on the human brain: From neurotoxicity, to vulnerability, to opportunity, *Front Neuroendocrinol*, 49, 91–105. *https://doi.org/10.1016/j.yfrne.2018.02.001*
19. Moncreiff, J., Cooper, R.E., Stockmann, T., Amendola, S., Hengartner, M.P. and Horowitz, A. (2022) The serotonin theory of depression: A systematic umbrella review of the evidence, *Mol Psychiatry. https://doi.org/10.1038/s41380-022-01661-0*
20. Campbell, C., Krugman, M., Lyons, T. and Miller, B. (2018) Factors that affect the stigma of mental illness in college students, *Mental Disorders Commons. http://hdl.handle.net/20.500.13013/639*
21. Link, B.G. (1987) Understanding labeling effects in the area of mental disorders: An assessment of the effects of the expectations of rejection, *Am Sociol Rev,* 52(1), 96–112. *https://psycnet.apa.org/doi/10.2307/2095395*
22. Vogel, L.D., Armstrong, P.I., Wade, N.G. and Tsai, P.C. (2013) Cross-sectional validity of the Self-Stigma of Seeking Help (SSOSH) scale: Examination across six nations, *J Couns Psychol,* 60(2). *https://doi.org/10.1037/a0032055*
23. Cornally, N. and McCarthy, G. (2011) Help-seeking behaviour: A concept analysis, *Int J Nurs Pract,* 17(3), 280–288. *https://doi.org/10.1111/j.1440-172X.2011.01936.x*
24. Kaier, E., Cromer, L.D., Johnson, M.D., Strunk, K. and Davis, J.L. (2015) Perceptions of mental illness stigma: Comparisons of athletes to nonathlete peers, *J Coll Stud Dev,* 56(7), 735–739. *http://doi.org/10.1353/csd.2015.0079*

25. Castaldelli-Maia, J.M., Gallinaro, J.G.D.M.E., Falcão, R.S., Gouttebarge, V., Hitchcock, M.E., Hainline, B., Reardon, C.L. and Stull, T. (2019) Mental health symptoms and disorders in elite athletes: A systematic review on cultural influencers and barriers to athletes seeking treatment, *Br J Sports Med,* 53, 707–721. *http://dx.doi.org/10.1136/bjsports-2019-100710*

26. Chapin, A. and McClure, K. (2020) An examination into the factors of stigma toward help seeking attitudes in college student athletes, *J Undergrad Res*, 25. *https://digitalcommons.lasalle.edu/undergraduateresearch/25*

27. Cutler, B.A. and Dwyer, B. (2020) Student-athlete perceptions of stress, support, and seeking mental health services, *J Issues Intercoll Athl,*13, 206–226. *http://csri-jiia.org/wp-content/uploads/2020/06/RA_2020_10.pdf*

28. Barnard, J.D. (2016) Student-athletes' perceptions of mental illness and attitudes toward help-seeking, *J College Stud Psychother,* 30(3), 161–175. *https://doi.org/10.1080/87568225.2016.1177421*

29. Bird, M.D., Chow, G.M. and Cooper, B.T. (2020) Student-athletes' mental health help-seeking experiences: A mixed methodological approach, *J College Stud Psychother,* 34(1), 59–77. *https://doi.org/10.1080/87568225.2018.1523699*

30. Vidourek, R.A., King, K.A., Nabors, L.A. and Merianos, A.L. (2014) Students' benefits and barriers to mental health help-seeking, *Health Psychol Behav Med,* 2(1), 1009–1022. *https://doi.org/10.1080/21642850.2014.963586*

31. Hillard, R.C., Redmond, L.A. and Watson II, J.C. (2019) The relationships among self-compassion, stigma, and attitudes toward counseling in student-athletes, *J Clin Sport Psychol,* 13(3), 374–389. *https://doi.org/10.1123/jcsp.2018-0027*

32. Tabet, S. (2019) An investigation of college student-athletes' mental health stigma, help-seeking attitudes, depression, anxiety, and life stress scores using structural equation modeling, *STARS: Showcase of Text, Archives, Research and Scholarship. https://doi.org/10.1080/19357397.2021.1924562*

33. O'Connor, M. and Casey, L. (2015) The Mental Health Literacy Scale (MHLS): A new scale-based measure of mental health literacy, *Psychiatry Res*, 229. https://doi.org/10.1016/j.psychres.2015.05.064

34. Fleming, C.J.E., Lynch, K.A., Hakas, M.B. and Belanger, E. (2018) Resource use after unwanted sexual experiences in undergraduates: A comprehensive evaluation of factors related to the decision to seek help, *J Appl Sport Psychol,* 1–20. *https://doi.org/10.1177/0886260518780408*

35. Lopez, R. (2008) The manifestation of depressed mood in student-athletes and their attitudes toward, barriers to, and preferences for seeking professional psychological help, *TRACE: Tennessee Research and Creative Exchange. https://trace.tennessee.edu/utk_graddiss/568*

36. Meng, X. and D'Arcy, C. (2015) Coping strategies and distress reduction in psychological well-being? A structural equation modelling analysis using a national population sample, *Epidemiol Psychiatr Sci,* 25(4), 370–383. *https://doi.org/10.1017/s2045796015000505*

37. Sullivan, P., Murphy, J. and Blacker, M. (2019) The level of mental health literacy among athletic staff in intercollegiate sport, *J Clin Sport Psychol,* 13(3), 440–450. *https://doi.org/10.1123/jcsp.2018-0052*

38. Wahto, R.S., Swift, J.K. and Whipple, L. (2016) The role of stigma and referral source in predicting college student-athletes' attitudes toward psychological help-seeking, *J Clin Sport Psychol,* 10(2), 85–98. *https://doi.org/10.1123/JCSP.2015-0025*

39. Biggin, I.R., Burns, J. H. and Uphill, M. (2017) An investigation of athletes' and coaches' perceptions of mental-ill health in elite athletes, *J Clin Sport Psychol,* 11(2), 126–147. *https://doi.org/10.1123/jcsp.2016-0017*

40. Kroshus, E. (2017) Stigma, coping skills, and psychological help seeking among collegiate athletes, *Athl Train Sports Health Care*, 9(6), 254–262. *https://doi.org/10.3928/19425864-20171010-02*

41. Anderson, A.J. and Dixon, M.A. (2019) How contextual factors influence athlete experiences of team cohesion: An in-depth exploration, *Eur Sport Manag Q*, 19(3), 353–372. *https://doi.org/10.1080/16184742.2018.1527381*

42. Chen, L.H., Kee, Y.H. and Chen, M.Y. (2015) Why Grateful Adolescent Athletes are More Satisfied with their Life: The Mediating Role of Perceived Team Cohesion, *Soc Indic Res,* 124, 463–476. *https://doi.org/10.1007/s11205-014-0798-0*

43. Delenardo, S. and Terrion, J. L. (2014) Suck it up: Opinions and attitudes about mental illness stigma and help-seeking behaviour of male varsity football players. *Can J Comm Ment Health*, 33(3), 43–56. *https://doi.org/10.7870/cjcmh-2014-023*

44. Conway, F.N., Domingues, M., Monaco, R., Lesnewich, L.M., Ray, A.E., Alderman, B.L, Todaro, S.M. and Buckman, J.F. (2020) Concussion symptom underreporting among incoming National Collegiate Athletic Association Division I college athletes, *Clin J Sport Med,* 30(3), 203–209. *https://doi.org/10.1097/JSM.0000000000000557*

45. Fogaca, J.L. (2021) Combining mental health and performance interventions: Coping and social support for student-athletes, *J Appl Sport Psychol,* 33(1), 4–19. *https://doi.org/10.1080/10413200.2019.1648326*

46. Hunt, J. and Eisenberg, D. (2010) Mental health problems and help-seeking behavior among college students, *J Adolesc Health,* 46(1), 3–10. *https://doi.org/10.1016/j.jadohealth.2009.08.008*

47. Kern, A., Heininger, W., Klueh, E., Salazar, S., Hansen, B., Meyer, T. and Eisenberg, D. (2017) Athletes-connected: Results from a pilot project to address knowledge and attitudes about mental health among college student-athletes, *J Clin Sport Psychol,* 11(4), 324–336. *https://doi.org/10.1123/JCSP.2016-0028*

48. Gulliver, A., Griffiths, K.M. and Christensen, H. (2012) Barriers and facilitators to mental health help-seeking for young elite athletes: A qualitative study, *BMC Psychiatry*, 12, 157. *https://doi.org/10.1186/1471-244X-12-157*

49. Evans, A.B., Blackwell, J., Dolan, P., Fahlen, J., Hoekman, R., Lenneis, V., McNarry, Smith, M. and Wilcock, L. (2020) Sport in the face of the COVID-19 pandemic: Towards an agenda for research in the sociology of sport, *Eur J Sport Soc,* 17(2), 85–95. *https://doi.org/10.1080/16138171.2020.1765100*

50. Yeo, T. J. (2020) Sport and exercise during and beyond the COVID-19 pandemic, *Eur J Prev Cardiol,* 27(12), 1239–1241. *https://doi.org/10.1177/2047487320933260*

51. Talevi, D., Socci, V., Carai, M., Carnaghi, G., Faleri, S., Trebbi, E., Bernardo, A.D., Cappelli, F. and Pacitti, F. (2020) Mental health outcomes of the CoViD-19 pandemic, *Riv Psichiatr*, 55(3), 137–144. *http://dx.doi.org/10.1708/3382.33569*

52. Graupensperger, S., Benson, A.J., Kilmer, J.R. and Evans, M.B. (2020) Social (un) distancing: Teammate interactions, athletic identity, and mental health of student-athletes during the COVID-19 pandemic, *J Adolesc Health,* 67(5), 662–670. *https://doi.org/10.1016/j.jadohealth.2020.08.001*

53. Sanborn, V., Todd, L., Schmetzer, H., Manitkul-Davis, N., Updegraff, J. and Gunstad, J. (2021) Prevalence of COVID-19 Anxiety in Division I Student-Athletes, *J Clin Sport Psychol*, 15(2), 162–176. *https://doi.org/10.1123/jcsp.2020-0057*

ABOUT STUDENT AUTHORS

Emma McCabe (she/her/hers) graduated with a bachelor's in exercise science with minors in biology and neuroscience from Elon University, NC in May 2021. During her undergraduate career, she performed concussion testing and original research with Dr. Caroline Ketcham and Dr. Eric Hall in Elon BrainCARE, exploring the relationship between nutrition and both mental and cognitive health. She is currently a postbaccalaureate research fellow at the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in the Human Psychopharmacology (HP) Lab under Dr. Vijay Ramchandani, studying determinants of alcohol use and misuse in clinical populations. Her postbaccalaureate research and manuscript under review (2022) examines latent class endorsement of positive coping behaviors during the COVID-19 pandemic and associations with alcohol-related and mental health outcomes. Emma's research interests are to learn motivations and barriers to health behaviors like physical activity, diet, and help-seeking behavior, and how health behaviors influence mental and physical health outcomes. She hopes this work will inform interventions that make positive lifestyle choices more accessible and sustainable for various populations.

Sarah DeSordi (she/her/hers) graduated from Elon University in 2021 with a BS in exercise science and dance science and a BFA in dance performance and choreography. Her undergraduate years were devoted to budding different paths of intersection between exercise science and dance. She did this by working as a part of the anatomy teaching assistant program, through nutrition and personal training internships, and pursuing a Pilates certification. Sarah continues to explore her interests in holistic wellness and human performance working as a Pilates instructor and freelance dancer, hoping to share her education and experiences with those she meets along the way.

PRESS SUMMARY

Student-athletes are more likely to develop mental health problems than the general population. Pressure to perform results in athletes playing through both physical injuries and mental health problems. Despite the high number of athletes who report needing mental health support, fewer than half seek out mental health services. Research has identified stigma as one of the largest barriers to mental health help-seeking behavior (HSB). Though stigma has been discussed in sports culture, few studies have focused on how different types of stigmas relate to HSB in student-athletes. The purpose of this study was to observe and describe student-athletes' perceived stigma (e.g., what others think) and personal stigma (e.g., what the individual thinks) in relation to HSB. N=20 athletes completed an online Qualtrics Survey, which included Link's Perceived Discrimination and Devaluation Scale, Mental Health Literacy Scale, Self-Stigma of Seeking Help Scale, Help Seeking Questionnaire, and Student-Athlete Role Behaviors Questionnaire. Results from this study may help develop mental health interventions to improve mental health HSB among student-athletes.

The Effect of Coastline Concavity on Maximum Storm Surge Height along the US Gulf Coast

Kayleigh Addington & Stephanie Zick*

Department of Geography, Virginia Tech, Blacksburg, VA

https://doi.org/10.33697/ajur.2023.084

*Student: kda99@vt.edu**
Mentor: sezick@vt.edu

ABSTRACT
Storm surge is the most dangerous component of landfalling tropical cyclones (TCs). The growing coastal population highlights the importance of research regarding the atmospheric and geographic factors influencing the maximum storm surge height (MSSH). To date, few studies have investigated the influence of coastline concavity. Here, we investigate the hypothesis that TCs making landfall on a concave coastline will have a higher MSSH than TCs making landfall on a convex coastline. The Colorado State University extended best track dataset includes the radius of 34 kt winds (R34), landfall minimum mean sea level pressure (MSLP), landfall maximum sustained winds, and forward speed of TCs. The storm surge database for the US Gulf Coast provides the location and MSSH for TCs impacting the U.S. Gulf Coast. From this, eleven TCs that meet specific criteria and represent the larger population of Atlantic TCs are selected. The adjusted degree of coastline concavity (ADoC) is calculated for each TC using the law of cosines and 50, 100, and 200 km radius buffers around the point of MSSH. A Mann Whitney *U* test does not indicate any significant differences between the mean MSSH of TCs making landfall on each coastline type. Additionally, results from a simple linear regression F-test suggest that none of the included parameters have a significant influence on MSSH despite the findings of previous research. Still, the Spearman's Rho correlation values suggest a weak positive relationship between the ADoC and MSSH. This relationship is significant at the 100 and 200 km buffers, which is consistent with the hypothesis. Results are limited by the small sample size. Future research should use a larger dataset and investigate how each individual storm characteristic affects MSSH.

KEYWORDS
Tropical Cyclones; Hurricanes; Storm Surge; Coastal Geography; Coastline Concavity; Gulf of Mexico; Law of Cosines

INTRODUCTION

With the increasing urbanization of coastal areas, the importance of local economies, and the growing number of people living near the ocean, research concerning coastal influences on storm surge is critical to developing effective coastal management, disaster preparedness, and resilience plans. The population density in counties along the U.S. Gulf Coast has grown by 32% since 1990.[1] Therefore, research focused on identifying locations of greater storm surge risk would allow emergency managers to formulate plans to expedite evacuation procedures for a growing population. For example, prior to the landfall of Hurricane Katrina (2005), many residents became stranded on evacuation routes due to the sheer number of people attempting to leave New Orleans at the same time.[2] This issue only contributed to the natural and social disaster that was Hurricane Katrina. Effective evacuation plans could streamline the evacuation process and decrease the amount of overcrowding on evacuation routes, allowing people to exit the area faster. Further knowledge of storm surge and the interactions between the coastline and maximum storm surge height (MSSH) has the potential to save lives.

Storm surge is often the most threatening and powerful component of landfalling tropical cyclones (TCs).[3] It can quickly reshape coastlines and destroy coastal habitats for both people and wildlife.[4] While TCs are ranked based on wind speed using the Saffir-Simpson Scale, storm surge is historically more destructive and deadly than wind. Storm surge height was once included in this scale but it was recently disassociated with the scale after events such as Hurricane Katrina (2005) and Hurricane Ike (2008) produced storm surges much higher than expected given the categorical rating of the storm at landfall.[4]

Numerous factors influence storm surge including the size and intensity (central barometric pressure) of a TC, the forward speed and angle at which the storm approaches the coastline, and the shape and size of the coastline and continental shelf.[5-10] The landfall maximum sustained winds, intensity, and size of a TC directly affect the storm surge height.[11] Sebastian *et al.* (2019) note that storm surge height on a concave coastline is dependent on the rate at which the water moves into the angled coast.[5] A fast-moving, intense TC with a high wind speed will result in a larger storm surge on open coasts due to a greater amount of water

pushed into the coastline by the wind and forward motion of the TC. However, a slow-moving TC causes water to be pushed into an enclosed coast for a longer time, which can result in a larger storm surge in these areas.[12] Additionally, a larger TC will have a higher storm surge because strong winds are affecting a larger area of water, forcing a greater volume of water into the coast. The approach angle of a TC also impacts storm surge height. Winds perpendicular to the coast at landfall have the most effect on storm surge.[6] This means TCs approaching from the south or southeast are more likely to have a greater storm surge due only to the orientation of the winds to the coastline. Conversely, Rogers and Davis (1993) show that TCs approaching the northern Gulf coast from the southeast will likely weaken more quickly,[13] and therefore have a smaller storm surge. More recent research shows that TCs tend to intensify as they approach land under favorable environmental circumstances.[14, 15] Intensification near landfall could also contribute to a greater storm surge since more intense TCs tend to produce larger storm surges. Sebastian *et al.* (2019) argue that the storm surge will be higher on a concave coastline due to the "convergence of energy and accumulation of a large volume of water into the coast" and will peak before landfall.[5] For example, Hope *et al.* (2013) note that Hurricane Ike produced a maximum surge in Chambers County, Texas, due to the county's coastline shape, bathymetry, and approach angle of the storm.[6] The northwest approach track and great size of Hurricane Ike combined with the gentle sloping continental shelf and concave coastline of this part of the Texas coast contributed to the production of a storm surge higher than expected given a category 2 ranking at landfall.[6]

Despite the more well-understood relationships described above, the characteristics of a TC and coastline often interact with each other and affect the storm surge height. Rogers and Davis (1993) show that TCs approaching a concave coastline experience a lower land-to-water ratio and, therefore, a faster pressure rise and subsequent weakening.[13] This could possibly result in a smaller storm surge despite the concave shape of the coastline. More recent research by Hope et al. (2013) and Sebastian *et al.* (2019) suggests otherwise. Additionally, Lok (2021) show that smaller TCs are more likely to intensify near landfall, and a more intense storm can result in a greater storm surge.[14] The intensification of TCs near the coastline can be linked to enhanced coastal downwelling.[15] Coastal downwelling keeps ocean waters warm near the coastline, which enhances convection and drives intensification despite the interaction of vertical wind shear and dry air near landfall.[15] This suggests that other factors, such as the curvature of the coastline, may affect the storm surge height.

This research will address the following question: Is there a relationship between the shape of the coastline (concave, convex) and the maximum height of storm surge in those areas? If so, do concave coastlines experience a higher storm surge than convex coastlines? Due to the claim by Sebastian *et al.* (2019) that concave coastlines experience an accumulation of water and energy,[5] we expect that storms making landfall on a concave coastline will have a higher maximum surge height than storms making landfall on a convex coastline.

METHODS AND PROCEDURES

The SURGEDAT dataset for landfalling TCs along the United States Gulf Coast, provided by Needham and Keim (2012),[4] is the primary database in this study. This dataset provides the height of the maximum storm surge in meters and location of maximum storm surge in latitude/longitude coordinates via measurements taken inside flooded buildings to minimize the effect of waves. The astronomical tide level as well as the influence of waves is removed from the reported MSSH for all events included in Needham and Keim (2012).[4] The storm surge events in Needham and Keim (2012) also have an associated confidence level (1-5) based on the type and credibility of the source reporting the MSSH.[4] A confidence level of 1 corresponds to low confidence and is used for events where there is only one information source from within 50 km of the location of maximum storm surge or when there are significant contradictions between sources. Conversely, a confidence level of 5 means the MSSH is supported by at least two credible sources and multiple tide gauges in the surrounding area.[4] **Figure 1** shows the distribution of storm surge heights across the Gulf Coast for the eleven storms included in the study.

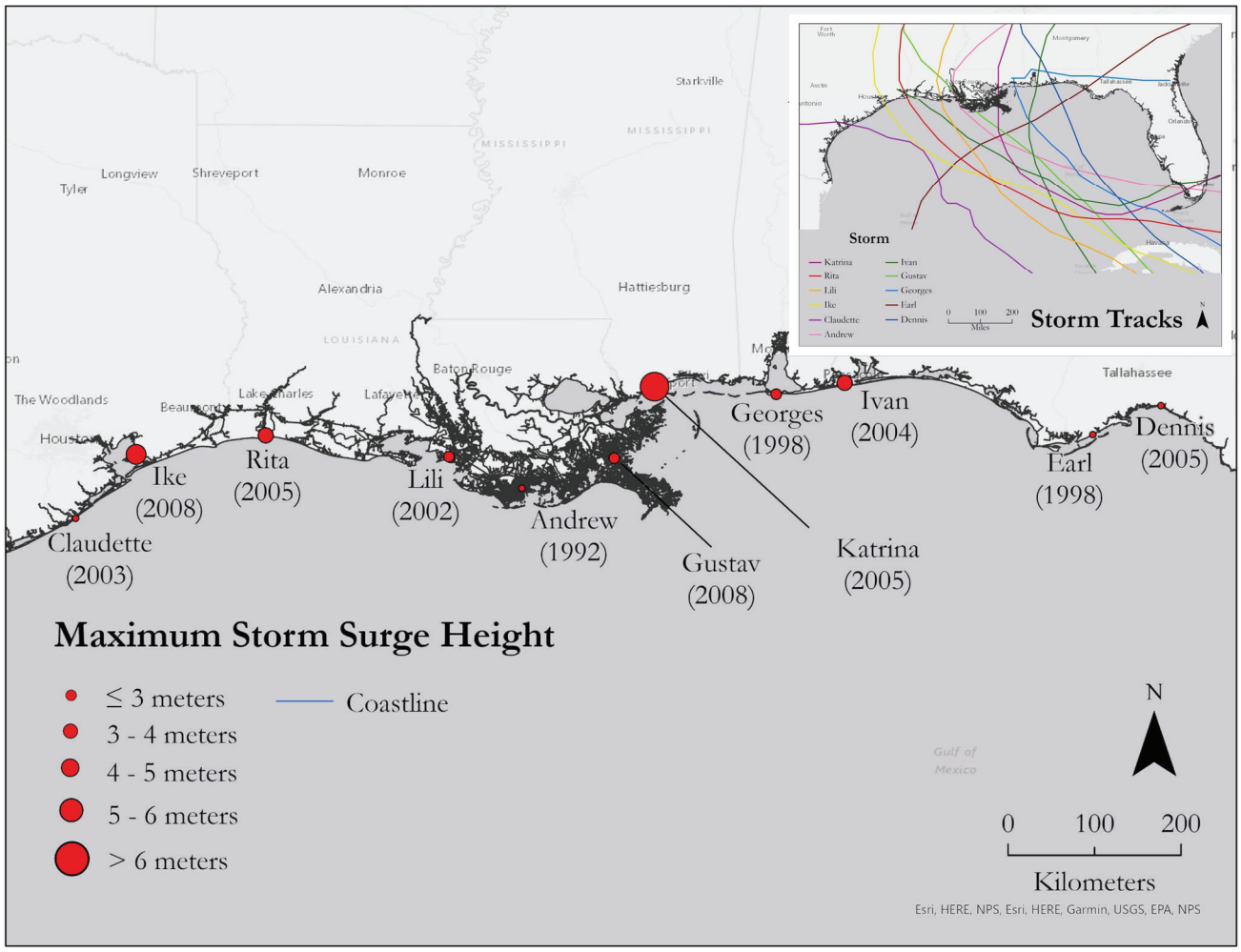


Figure 1. Geographical overview of the location and height of each maximum storm surge associated with the eleven TCs in this study. The inset shows the tracks of each TC.

This research focuses on TCs with a storm surge greater than 2 meters and a ranking of category 1 (64 knots) or higher on the Saffir-Simpson Hurricane Wind Scale. TCs in this study have landfall wind speeds between 75 and 125 knots, centered at 100 knots. Additionally, for a storm to be included in the analysis, it must meet the following criteria. First, the MSSH measurement must have a confidence level of 3 or higher (moderate to very high confidence). This means the MSSH must be validated by at least one credible source that is not contradicted by any other source.[4] Additionally, Drake (2012) classifies TCs with R34 measurements between 126 and 174 nautical miles as "medium" size. Therefore, TCs in this study must have an R34 measurement between 100 and 200 nautical miles with the distribution centered around 150 nautical miles.[16] Based on data from the HURDAT2 dataset, the average MSLP for TCs in the Atlantic basin occurring between 1988 and 2012 is 955 hPa.[17] The TCs in this study have an average landfall MSLP of 950 hPa. Based on data from the National Oceanic and Atmospheric Administration (NOAA), the forward speed of TCs between 5 and 30 degrees north latitude is between 10-14 mph.[18] The average forward speed for TCs in this study is 12 mph. Lastly, all TCs in this study make landfall along the northern Gulf coast so the angle of incidence with the coastline is visually between 60 to 120 degrees and all TCs have a northerly component to the forward motion. Specific measurement of the angle of approach is outside the scope of this research, which focuses on the coastline curvature, though it should be considered in future work.

The extended best track database, obtained from Colorado State University and described in Demuth (2006), is used to determine the landfall wind speed, pressure, and radius of 34-knot winds (R34) for each storm.[19] Additional specific statistics on individual storms are obtained from local National Weather Service websites and the National Hurricane Center website.[20-37] Since the sample of TCs in this research is centered around a medium size and average pressure and forward speed, these storms generally reflect the larger distribution of all hurricanes in the Atlantic basin. Summary statistics, including the mean and median for each storm attribute are shown in **Table 1**.

After selecting storms that meet the criteria listed above, we map the coordinates and determine the degree of concavity of the surrounding coastline. The NOAA Shoreline Website (NOAA 2000, 2008, 2013) provides the shapefile containing the coastline mapping data for the United States. The shoreline dataset is created using NOAA nautical charts and the mean high water tidal datum. The shapefile has an average scale of 1:70,000.[38] This file is clipped to focus only on the Gulf Coast and later to locations surrounding each maximum storm surge point.

In their research concerning the SURGEDAT database, Needham and Keim (2012) include tables with storm surge heights for numerous systems.[4] This database defines the spatial domain of this study as the United States Gulf Coast. The Gulf Coast provides a similar bathymetric profile at nearly all landfall locations,[6, 39, 40] thus eliminating one major influence on storm surge height and making it an ideal location to compare the storm surge height of different storms. The amount of available data included in this dataset and in the extended best track database restricts the temporal domain to events occurring between 1988 and 2012.[4, 19]

Storm Name	R34 (nautical miles)	Landfall Minimum MSLP (hPa)	Landfall Maximum Sustained Winds (kts)	Forward Speed (mph)
Andrew (1992)	125	937	125	16
Earl (1998)	103.75	985	80	10
Georges (1998)	121.25	961	95	7
Lili (2002)	157.5	957	105	15
Claudette (2003)	103.75	982	75	12
Ivan (2004)	187.5	931	110	12
Dennis (2005)	145	930	120	18
Rita (2005)	145	931	105	11
Gustav (2008)	180	954	95	16
Ike (2008)	190	952	95	10
Katrina (2005)	162.5	923	110	15
Mean	**144.26**	**949.15**	**100.26**	**12.49**
Median	**145.00**	**952.00**	**105.00**	**12.00**

Table 1. Characteristics of the storms used in this study.

After storms meeting all criteria are selected and mapped, we assign a value of concavity to the landfall area for each storm based on the section of coastline surrounding a maximum storm surge point. **Figure 2** explains the calculation of the degree of concavity at a 50 km radius around the point of maximum storm surge. After we map the location of each maximum storm surge, we create 50, 100, and 200 km buffers around each point of maximum storm surge.

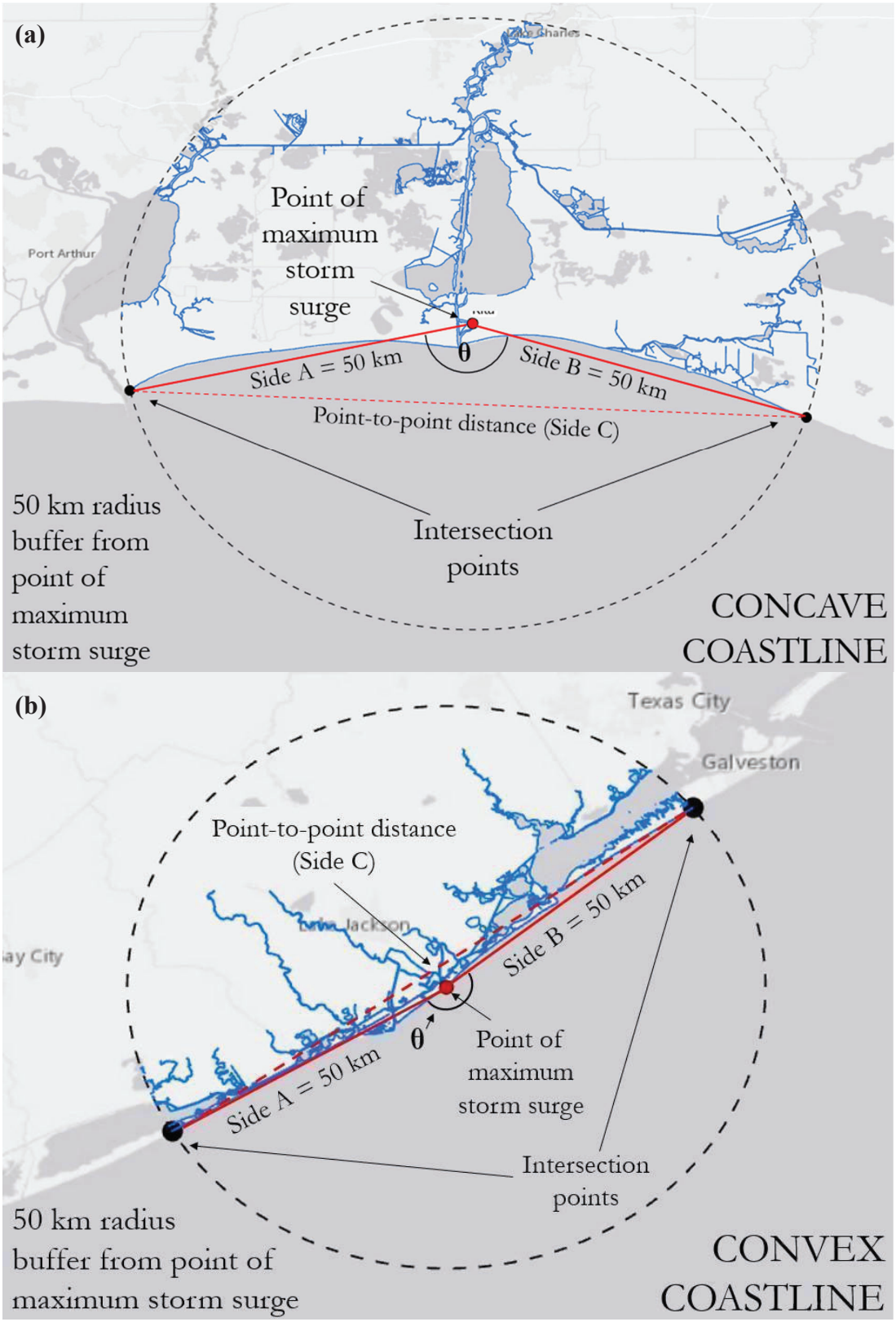


Figure 2. Calculation of the degree of concavity on a (a) concave and (b) convex coastline.

Next, we locate the intersections of this buffer with the coastline (**Figure 2**). To determine the degree of concavity, we calculate the straight-line distance between the intersection points (**Figure 2**). Then, the Law of Cosines is used to calculate the angle (θ_{rad}) between the coastline-buffer intersection points at radii of 50, 100, and 200 km. Sides A and B are equal to the size of the radius. For example, a radius of 50 km (100 km, 200 km) means sides A and B are equal to 50 km (100 km, 200 km). The point-to-point distance is represented by side C (**Table 2**). This process results in an angle measurement in radians (θ_{rad}), which is then converted to degrees (θ). Using this angle measurement in degrees, the adjusted degree of concavity (ADoC) is calculated using **Equation 1**:

$$\text{ADoC} = \frac{180-\theta}{180} \qquad \textbf{Equation 1.}$$

A coastline is concave if side C is located seaward of the point of maximum storm surge. In this situation, θ is less than 180 degrees and the ADoC is greater than zero. A coastline is convex if side C is located landward of the point of maximum storm surge. Here, θ is greater than 180 degrees and the ADoC is less than zero. A perfectly straight coastline has an angle of 180 degrees and an ADoC equal to zero. It is important to note that the coastline becomes more concave with larger buffer sizes due to the concave nature of the U.S. Gulf Coast. Additionally, some points of maximum storm surge are not located directly on the coastline since the original measurements were taken from inside flooded buildings.[4] This results in some error in calculating the ADoC because the location of the buffer-coastline intersection may cause the distance between intersection points to be slightly different than if the point of maximum storm surge were located directly on the coastline. In this study, the difference is less than 1% for MSSH points located on open coastlines but can range as high as 33% for MSSH points located near bays and deltas. In all cases, the category of concave or convex is unchanged. Another source of error could be due to the resolution of the NOAA shoreline dataset. Slight differences between this dataset and the actual coastline could affect the location of the intersection points, the length of side C, and therefore, the calculation of the ADoC.

In this study, the sample size is small, and the data are not normally distributed (see results), so non-parametric tests are necessary to evaluate the significance of the results. For all statistical tests, we use a 95% confidence level ($p < 0.05$) to determine the significance of the result. First, the Spearman's Rho is calculated to determine the correlation between MSSH and the ADoC for each buffer size. Second, a Mann Whitney *U* test is used to evaluate whether significant differences exist between MSSH associated with concave and convex coastlines. In general, the Mann Whitney *U* test determines if the samples come from different populations or if they are rooted in the same population. As a non-parametric test, the data do not have to be normally distributed since this test focuses on the rank of the data rather than the actual values. Still, this test has low power with small sample size and is only applicable when groups have four or more data points. In fact, it is impossible to achieve a statistically significant result ($p < 0.05$) with less than four samples in a group.[41] For this reason, the Mann Whitney *U* test is only used to compare the concave and convex groups at the 100 km buffer. Despite these limitations, the Mann Whitney *U* test is the best statistical test option to compare the data and serves as a model for future work that includes more cases. The null hypothesis (h0) is that there is no difference between the average storm surge height on a concave coastline and a convex coastline. The alternative hypothesis (h1) is that there is a difference between the average storm surge height on a concave coastline and a convex coastline.

Next, we perform a simple linear regression F-test. This test investigates the linear relationship between MSSH and each individual storm characteristic by comparing the sum of squares error of the full model to the sum of squares error of a reduced model. In the context of this research, the full model includes a linear relationship between an individual storm characteristic and the MSSH. Each reduced model contains only the y-intercept and residual error for that specific storm characteristic. There are no other variables in the reduced model suggesting no relationship, which is the null hypothesis. The F statistic tells us if the full model better explains any relationship (if it exists) between an individual storm characteristic and storm surge height. A large F-statistic with a significant ($p < 0.05$) result indicates that we should reject the null hypothesis for that individual storm characteristic and suggests that a specific variable has a significant influence on storm surge height.[42] Here, we investigate seven full models comparing, individually, the R34, landfall minimum MSLP, landfall maximum sustained winds, forward speed, ADoC at a 50 km radius, ADoC at a 100 km radius, and ADoC at a 200 km radius to the MSSH.

50 km Radius Buffer					
Hurricane	**MSSH (m)**	**Coastline Type**	**Side C**	**Angle (deg)**	**ADoC**
Ike	5.33	Concave	88.58	124.7	0.307
Ivan	4.57	Concave	99.72	171.4	0.047
Rita	4.57	Concave	96.53	149.7	0.168
Dennis	2.74	Concave	85.94	118.5	0.342
Earl	2.44	Convex	98.40	200.5	-0.114
Georges	3.63	Convex	98.03	202.8	-0.127
Gustav	3.96	Concave	57.07	69.6	0.613
Lili	3.75	Concave	86.75	120.3	0.331
Claudette	2.79	Convex	99.69	189.0	-0.050
Andrew	2.44	Concave	98.40	159.5	0.114
Katrina	8.47	Concave	97.78	155.8	0.134

100 km Radius Buffer					
Hurricane	**MSSH (m)**	**Coastline Type**	**Side C**	**Angle (deg)**	**ADoC**
Ike	5.33	Concave	188.07	140.2	0.221
Ivan	4.57	Convex	199.78	185.4	-0.030
Rita	4.57	Concave	192.93	149.4	0.170
Dennis	2.74	Concave	156.93	103.4	0.426
Earl	2.44	Convex	194.00	208.1	-0.156
Georges	3.63	Convex	197.31	198.8	-0.105
Gustav	3.96	Convex	190.61	215.3	-0.196
Lili	3.75	Concave	197.84	163.2	0.094
Claudette	2.79	Convex	199.06	191.1	-0.062
Andrew	2.44	Convex	194.33	207.4	-0.152
Katrina	8.47	Concave	153.03	99.8	0.445

200 km Radius Buffer					
Hurricane	**MSSH (m)**	**Coastline Type**	**Side C**	**Angle (deg)**	**ADoC**
Ike	5.33	Concave	377.36	141.3	0.215
Ivan	4.57	Concave	375.82	140.0	0.222
Rita	4.57	Concave	375.75	139.9	0.223
Dennis	2.74	Concave	372.08	136.9	0.239
Earl	2.44	Concave	399.97	178.6	0.008
Georges	3.63	Concave	387.48	151.3	0.160
Gustav	3.96	Convex	393.49	200.7	-0.115
Lili	3.75	Concave	399.22	172.9	0.040
Claudette	2.79	Concave	399.51	174.3	0.031
Andrew	2.44	Convex	361.05	231.0	-0.283
Katrina	8.47	Concave	372.20	137.0	0.239

Table 2. Calculation of the adjusted degree of concavity for each storm at a 50, 100, and 200 km radius.

RESULTS

Since storm surge height can be affected by the intensity, forward speed, size, and landfall wind speed, this study uses sample TCs with characteristics reflecting the larger population of TC landfall characteristics (**Table 1**). All TCs included have an average landfall wind speed of roughly 100 kts. These TCs have an average forward speed of 10.9 knots, an average landfall pressure of 949.2 hPa, and an average R34 of 144.3 nautical miles. The median values for each of these parameters are similar to the mean values (**Table 1**). The values for landfall maximum sustained winds are normally distributed (not shown); the values for R34, landfall minimum MSLP, and forward speed are not normally distributed. While the distributions of R34, landfall minimum MSLP, and forward speed values are not normal, the TCs selected still reflect a larger population of R34, landfall minimum MSLP, and forward speed values for TCs in the North Atlantic basin.[16-19]

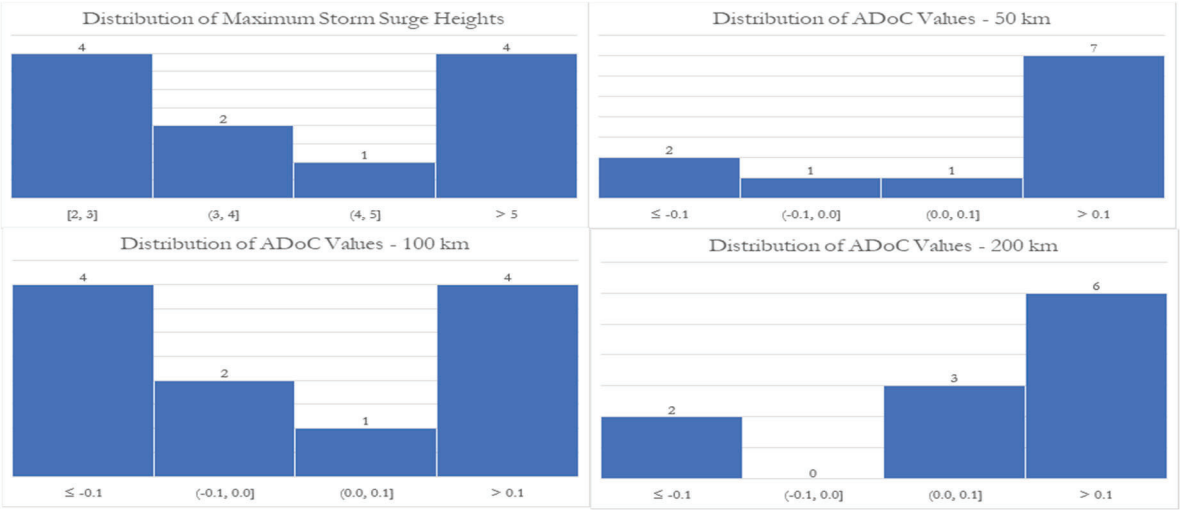


Figure 3. Histograms showing the distribution of storm surge heights (m) data and degrees of concavity.

Figure 3 shows the distribution of MSSH and ADoC at each buffer size. The ADoC values are not normally distributed, so non-parametric tests, such as Spearman's Rho and the Mann Whitney U test, are needed for analysis. Upon investigation, there is a slight correlation (r_s=0.292) between the ADoC and the MSSH on a 50 km radius buffer. The correlation increases to r_s =0.552 on a 100 km radius buffer and to r_s =0.525 on a 200 km radius buffer (**Table 3**). Using a 95% confidence level, the 100 and 200 km correlations are significant. **Figure 4** shows a visual representation of the relationship between MSSH and the ADoC at each buffer size. Just as the Spearman's Rho correlations indicate a weak relationship between variables, **Figure 4** suggests a slight positive relationship between the MSSH and the ADoC. Further, the relationship between these two variables strengthens as the buffer size increases, which is also supported by the significant relationships at the 100 and 200 km buffers. Given the numerous factors affecting storm surge height, a positive relationship for all three buffer sizes is notable, especially since the relationship is consistent with prior research suggesting higher storm surges along concave coastlines.[5]

Storm Name	Rank MSSH	Rank 50 km ADoC	Rank 100 km ADoC	Rank 200 km ADoC
Earl	1.5	2	2	3
Andrew	1.5	5	3	1
Dennis	3	10	10	11
Claudette	4	3	5	4
Georges	5	1	4	6
Lili	6	9	7	5
Gustav	7	11	1	2
Ivan	8.5	4	6	8
Rita	8.5	7	8	9
Ike	10	8	9	7
Katrina	11	6	11	10
Spearman's Rho (*p*-value)		**0.292 (0.192)**	**0.553 (0.039)**	**0.525 (0.049)**

Table 3. Ranked value for each variable used in the calculation of Spearman's Rho at each buffer size. The Spearman's Rho value is also shown with the corresponding *p*-value.

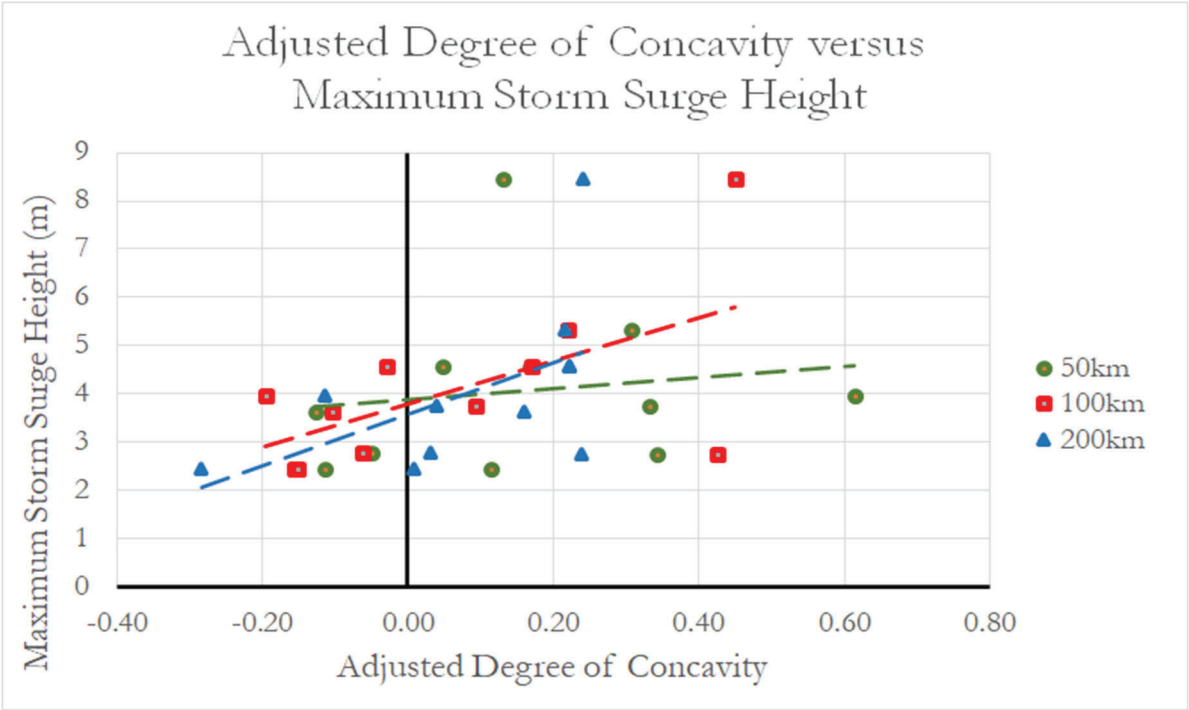


Figure 4. Scatter plot showing MSSH versus adjusted degree of concavity. Dashed lines indicate the approximate linear relationship for each buffer size.

Figure 5 compares the average MSSH for TCs making landfall on a concave coastline to the average MSSH of TCs making landfall on a convex coastline at each buffer size. Using the 50 km radius buffer, eight storms make landfall on a concave coastline, while three storms make landfall on a convex coastline. At a 100 km radius buffer five storms make landfall on a concave coastline and six storms make landfall on a convex coastline. Using the 200 km radius buffer, nine storms make landfall on a concave coastline while two storms make landfall on a convex coastline. For each buffer size, the average storm surge height on a concave coastline is higher than the average storm surge height on a convex coastline. The Mann Whitney *U* test is completed by comparing two groups (concave vs. convex coastlines) for the 100 km buffer only using a 95% confidence level. This test is not used to compare concave and convex groups at the 50 and 200 km buffers since there are less than four data points in a single group at each of these buffer sizes. The *p*-value (**Table 4**) does not indicate a statistically significant difference between the median height of maximum storm surge on a concave coastline and median height of maximum storm surge on a convex coastline at the 100 km buffer size. However, we must be cautious of these results due to the small sample size. Still, the results indicate that the null hypothesis may not be rejected and that storm surges in the two groups are statistically similar.

Buffer Size	*p*-value
100 km	0.118

Table 4. Mann Whitney *U* test *p*-values for each buffer size.

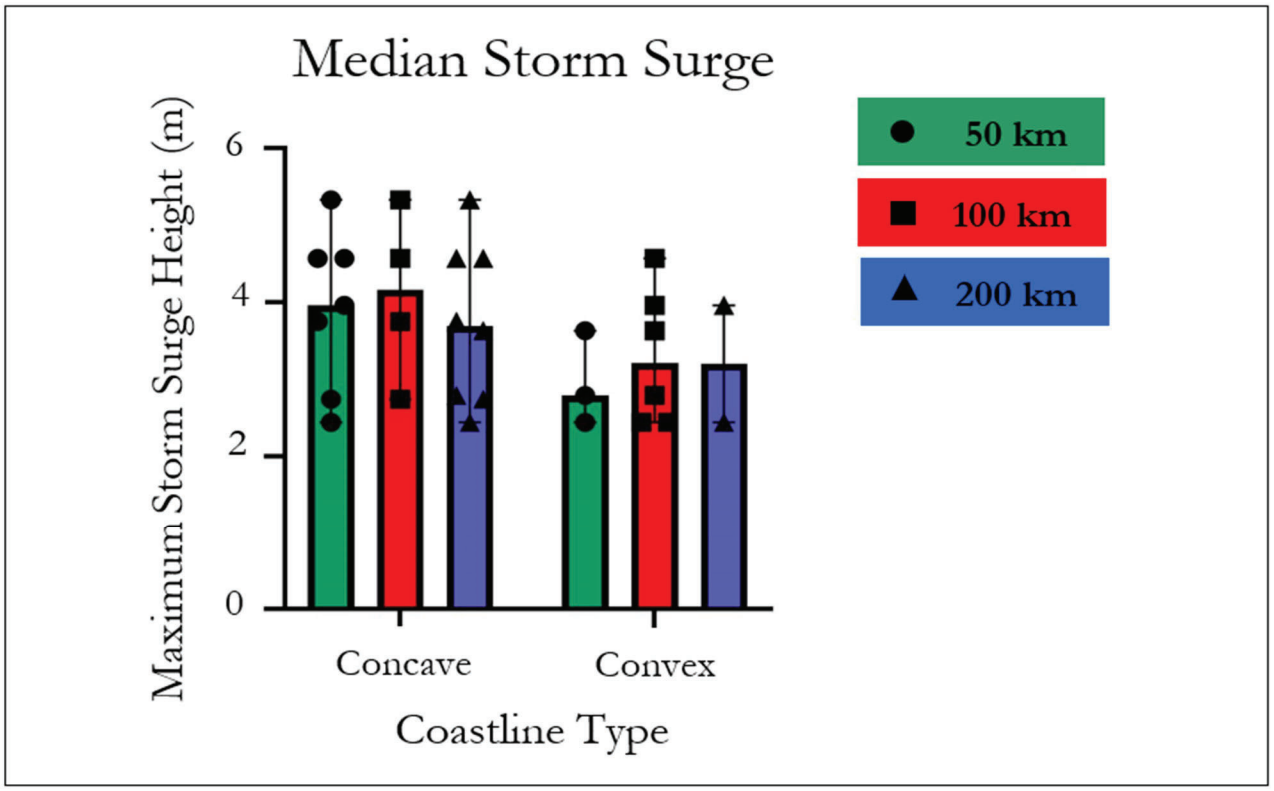


Figure 5: Comparison of the distributions of MSSH along a concave coastline and a convex coastline at each radius. The bar represents the median MSSH on each type of coastline. The whiskers represent the range. The points represent individual MSSH in the respective groups.

To investigate the relationship between MSSH and other individual storm characteristics, an F-test is performed (**Table 5**). First, to ensure linear relationships between all variables and the MSSH, we use a log transformation for the MSLP variable. All other variables have a near-linear relationship with MSSH and are not manipulated. For this sample, F-statistics indicate that there are no statistically significant influences on storm surge height at a 95 % confidence level (**Table 5**) despite the findings of previous research suggesting that R34, the landfall maximum sustained winds, landfall minimum MSLP, forward speed, all influence the MSSH.[5-15] For example, the explanatory power of the transformed landfall minimum MSLP [log (landfall minimum MSLP)] is not significant in this research, but previous studies show that MSLP does have a well-established relationship with storm surge height.[5, 7, 43, 44]. In this sample, multiple variables with established relationships with storm surge do not show significant linear relationships, which may be due to the small sample size. Therefore, we interpret these results cautiously and do not rule out the influence of any variable on the MSSH.

Despite the small sample size, we attempt to build a multiple linear regression model using forward stepwise regression since many of the variables achieve a 10% significance threshold ($p < 0.1$). This model tests if the variables included in Table 5 significantly predict the MSSH and uses the Akaike Information Criteria (AICc) stopping rule, similar to other studies in the atmospheric and climate science fields.[45, 46] The landfall maximum sustained winds, log (landfall minimum MSLP), R34, forward speed, and ADoC at each buffer size is considered in a single model. Using the AICc stopping rule, only the ADoC at the 100 km buffer is recommended for the model ($F = 4.777$; p-value = 0.057). This result implies that a simple linear regression model is the best linear model to predict MSSH and that additional variables do not explain enough additional variance to be included in the final model. However, as stated earlier, multiple variables with established relationships with MSSH are not included in the model, so results should be interpreted cautiously.

Parameter	F-Ratio (*p*-value)
Landfall Maximum Sustained Winds (kts)	0.246 (0.632)
log (Landfall Minimum MSLP (hPa))	3.433 (0.097))
R34 (nautical miles)	4.270 (0.069)
Forward Speed (mph)	0.037 (0.851)
ADoC - 50 km	0.216 (0.653))
ADoC - 100 km	4.777 (0.057)
ADoC - 200 km	3.492 (0.095)

Table 5. Simple linear regression F-ratio and corresponding *p*-value for each parameter, including the ADoC at each buffer size.

DISCUSSION

A handful of previous research studies address the many factors affecting storm surge height, but few studies have investigated the effect of coastline concavity, alone, on storm surge height. This study attempts to limit the influence of outlier storms by using a sample of TCs that generally reflects the larger population of Atlantic TCs in terms of intensity, size, landfall wind speed, and forward speed values. All TCs in this study make landfall along the US Gulf Coast, so the influence of the approach angle and the bathymetry of the continental shelf is also limited. Based on findings from previous research, it is hypothesized that landfalling TCs will have a higher MSSH on a concave coastline than similar landfalling TCs with MSSH on a convex coastline.

The Spearman's Rho values suggest a weak correlation between the MSSH and ADoC at the three radii used. Given the small sample size and considering the numerous factors affecting storm surge height, these correlations are a compelling result of this study, especially since the 100 and 200 km correlations are statistically significant. Additionally, a similar and increasingly positive trend with increasing buffer radius is an interesting result, further supporting the relationship (**Figure 5**). While the Mann Whitney *U* test does not suggest a statistically significant difference between the MSSH on a concave coastline and the MSSH on a convex coastline at the 100 km buffer, the sample size might be too small to detect any differences in the groups. Furthermore, acknowledging the numerous factors affecting storm surge height, it might be difficult to detect a clear relationship solely between the degree of coastline concavity and MSSH. Additionally, a forward stepwise regression suggests that only the ADoC at the 100 km buffer has significant explanatory power and is the only variable included in the final linear regression model. With a larger dataset, a multiple linear regression could be used with more confidence to evaluate whether coastline curvature can explain additional variation after other primary factors such as storm size are first considered.

Given the weak correlation between storm surge height and coastline curvature, it is interesting to note that, using a 50 km radius, the greatest storm surge observed on a convex coastline is greater than only two storm surges observed on a concave coastline. This trend is not as well pronounced with the 100 km or 200 km radius buffers (**Figure 5**). However, the median MSSH on convex coastlines is consistently smaller than the median MSSH on concave coastlines at each radius (**Figure 5**). These differences are small and not statistically significant based on a Mann Whitney *U* test. The small sample size could affect the outcome of the Mann Whitney *U* test. Nevertheless, combined with the increasing correlation between the ADoC and the MSSH as the buffer radius increases, which reaches significance at the 100 and 200 km buffers, these observations suggest the potential for even stronger results when a larger pool of storms is considered. Additionally, because the correlation between the ADoC and the MSSH increases and becomes significant with increasing buffer size (**Table 3, Figure 5**), a larger buffer zone size may be a more appropriate scale for assessing the influence of coastline shape. Future research should investigate the appropriate horizontal scale for assessing the influence of coastline curvature on storm surge.

There are limited studies investigating the effect of coastline concavity on MSSH, but the results of this study match the expectations set by the conclusions of Sebastian *et al.* (2019)[5] and Hope *et al.* (2013).[6] In their research, Sebastian *et al.* (2019) show that concave coastlines will likely experience a higher storm surge due to the concentration of water into the inward-sloping coast; the results of this study suggest similar findings in that most storm surges on a concave coastline are higher than the storm surges occurring on a convex coastline.[5] Additionally, Hope *et al.* (2013)[6] note that the concave coastline in Louisiana and Texas contributed to the higher storm surge in Hurricane Ike.[6] The results of this study provide further evidence that a positive

relationship exists between coastline curvature and storm surge height, suggesting that concave coastlines may contribute to increased storm surge height.

The primary source of error in this study stems from the small sample size. In particular, the amount of data available in both databases used in this study restricts the temporal range from 1988-2012 and introduces further sources of error due to insufficient data available for meaningful statistical significance tests. Additionally, to limit the influence of the known factors affecting storm surge height, this study only includes storms with landfall characteristics that meet certain criteria and reflect a larger population of Atlantic TCs, which further limits the sample size. Due to the small sample size and non-normal distributions of variables, this study is limited in the statistical tests that can be used. For example, while the F-test did not identify any significant variables, we attempt to build a simple linear regression model, which only recommends the use of the ADoC at the 100 km buffer in the final model despite established relationships between MSSH and other variables. Furthermore, the Mann Whitney *U* test is only applicable at the 100 km buffer, and the results may not accurately reflect any true differences in the population due to the small sample size. As a result, there is an opportunity to conduct further research investigating the effect of coastline curvature when larger geographical and temporal scales are considered. Despite the limitations in the dataset, the approach used in this study may still be useful in future studies that seek to separate out the impact of a specific storm characteristic (*e.g.* intensity, wind speed, size, or forward speed) on MSSH. Changes in the coastline shape over time as the result of sea level rise and erosion/mitigation efforts and the influence of coastal waterways (rivers, bays, *etc.*) may also impact the degree of concavity and/or MSSH in an area.

An additional source of error stems from the location of MSSH with respect to the coastline. Some points of maximum storm surge are not located directly on the coast as these measurements were taken from inside flooded buildings.[4] Because the points of intersection between the coastline and each buffer are located directly on the coast, the distance of maximum storm surge from the coast has the potential to affect the angle measurement used in the calculation of the degree of concavity in this study, especially for the 50 km buffer. However, this error is generally small since the buildings are usually within 1 km of the coastline. To further minimize this effect, the 100 km and 200 km buffer zones are included. However, as the buffer zone is increased, the small inflections along each coastline are minimized and the degree of concavity of the entire section of coastline could be affected, resulting in an "average" degree of coastline concavity. More specifically, using a smaller buffer size allows small changes in the coastline shape to be better reflected in the ADoC calculation. For example, the 50 km buffer better captures the coastline shape of small bays or inlets around the point of maximum storm surge, which can affect storm surge height.[48,49] The larger buffer size (100 km or 200 km) covers areas of the coastline that may include bays or inlets but does not account for the coastline shape within these bays and inlets. In other words, a larger buffer does not capture these small features as well as a smaller buffer so the ADoC value becomes more of an "average" for the whole area. Idealized modeling studies could be used to eliminate this error by obtaining MSSH values directly at the coastline.

CONCLUSIONS

This research focuses on the effect of coastline concavity on MSSH along the Gulf Coast of the United States. More specifically, this research addresses the following questions: Is there a relationship between the shape of the coastline (concave, convex) and the maximum height of storm surge in those areas? If so, do concave coastlines experience a higher storm surge than convex coastlines? While there are few studies investigating the effect of coastline concavity on storm surge height, many studies state that storm surge height is influenced by a number of factors. The intensity (central barometric pressure) of a cyclone, forward speed and angle at which the storm approaches the coastline, and the shape and size of the coastline and continental shelf all influence storm surge height.[5-11] Many of these factors can also interact to influence storm surge height. Because of this, this study only uses TCs that meet certain criteria to more closely investigate the effect of coastline concavity on storm surge height. These criteria ensure these TCs represent the greater population of TCs with respect to size, intensity, landfall wind speed, and forward speed values. This study also focuses on the Gulf Coast of the United States to limit the influence of bathymetry and approach angle of TCs.

Results from a Mann Whitney *U* test did not indicate any significant difference between the storm surge heights of TCs making landfall on concave coastlines and the storm surge heights of TCs making landfall on convex coastlines at the 100 km buffer. This test was not applicable at the 50 and 200 km buffers due to the small number of samples in each group. The small sample size could limit the accuracy of this test at the 100 km buffer as well. Additionally, a simple linear regression F-test shows there are no significant influences of MSSH. This result is also likely due to the small sample size since many of these influences are nearly significant and previous research describes established relationships between storm size, storm speed, and storm intensity. For this reason, forward stepwise regression is used and shows that ADoC at the 100 km buffer is the dominant influence on MSSH. These results should be interpreted with caution due to the small sample size and absence of significant ($p < 0.05$) F-ratios prior to building the simple linear regression model. The Spearman's Rho correlation values suggest a weak positive relationship

between the coastline concavity and the MSSH. This correlation is significant when considering concavity at larger horizontal scales (100 and 200 km). Additionally, storm surges on concave coastlines are consistently greater than storm surges on convex coastlines.

As mentioned above, a small dataset (n=11) marks the greatest limitation of this research. Expanding the spatial and temporal range of this research to include more TCs affecting the United States Gulf Coast and TCs impacting the southeast-Atlantic and northeast-Atlantic coastlines would add additional data to the analysis and possibly reveal additional significant relationships. Further research into the effect of individual storm characteristics on storm surge height would also provide more context and a better understanding of the potential impact of TCs on coastal communities. Lastly, this study is observation-based. Modeling studies and case studies would be helpful in isolating and identifying the influence of coastline curvature on storm surge height.

REFERENCES

1. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Storm Surge Overview, *https://www.nhc.noaa.gov/surge/* (Accessed Nov 2020)
2. Debrovo, A. (2008) *Hurricane Katrina - The One We Feared*, 14, Express Publishing, Harahan, LA.
3. Rappaport, E. N., 2014: Fatalities in the United States from Atlantic Tropical Cyclones: New Data and Interpretation. *Bulletin of the American Meteorological Society*, **95**, 341–346, *https://doi.org/10.1175/BAMS-D-12-00074.1*.
4. Needham H. F. and Keim B. D. (2012) A storm surge database for the US Gulf Coast, *Int J Climatol,* **32**, 2108-2123. *https://doi.org/10.1002/joc.2425*
5. Sebastian M., Behera M. R., and Murty P. L. N. (2019) Storm Surge Hydrodynamics at a Concave Coast due to varying approach angles of cyclone, *Ocean Eng,* **191**, 1-16. *https://doi.org/10.1016/j.oceaneng.2019.106437*
6. Hope, M. E., Westerink, J. J., Kennedy, A. B., Kerr, P. C., Dietrich, J. C., Dawson, C., Bender, C. J., Smith, J. M., Jensen, R. E., Zijlema, M., Holthuijsen, L. H., Luettich Jr., R.A., Powell, M. D., Cardone, V. J., Cox, A. T., Pourtaheri, H., Roberts, H. J., Atkinson, J. H., Tanaka, S., Westerink, H. J., Westerink, L. G. (2013) Hindcast and Validation of Hurricane Ike (2008) waves, forerunner, and storm surge, *J. Geophys. Res. Oceans,* **118**, 4424-4460. *https://doi.org/10.1002/jgrc.20314*
7. Resio, D. T. and Westerlink, J. J. (2008) Modeling the physics of storm surges, *Physics Today,* **61**, 33-38. *https://doi.org/10.1063/1.2982120https://doi.org/10.1063/1.2982120*
8. Irish, J. L., D. T. Resio, and J. J. Ratcliff, (2008) The Influence of Storm Size on Hurricane Surge. *Journal of Physical Oceanography*, **38**, 2003–2013, https://doi.org/10.1175/2008JPO3727.1.*https://doi.org/10.1175/2008JPO3727.1*.
9. Weisberg, R. H., and L. Zheng, (2006) Hurricane Storm Surge Simulations for Tampa Bay. *Estuaries and Coasts*, **29**, 899–913. *https://www.jstor.org/stable/4124819*
10. Zhang, C., and C. Li, (2019) Effects of hurricane forward speed and approach angle on storm surges: an idealized numerical experiment. *Acta Oceanol. Sin.*, **38**, 48–56, *https://doi.org/10.1007/s13131-018-1081-z*.
11. Longshore, D. (2008, 2000) *Encyclopedia of Hurricanes, Typhoons, and Cyclones* New Edition, 380-384, Checkmark Books, New York.
12. Islam, R., and H. Takagi, (2021) Statistical significance of tropical cyclone forward speed on storm surge generation: retrospective analysis of best track and tidal data in Japan. Georisk: *Assessment and Management of Risk for Engineered Systems and Geohazards* **15**, 247–257, https://doi.org/10.1080/17499518.2020.1756345.
13. Rogers R. F. and Davis R. E. (1993) The Effect of Coastline Curvature on the Weakening of Atlantic Tropical Cyclones, *Int J Climato,l* **13**, 287-299. *https://doi.org/10.1002/joc.3370130305*
14. *Lok, C. C. F., Chan, J. C. L., Toumi, R. (*2021*)* Tropical cyclones near landfall can induce their own intensification through feedbacks on radiative forcing, *Communications Earth and Environment,* **2:184***,* 1-10. *https://doi.org/10.1038/s43247-021-00259-8*.
15. Gramer, L. J., Zhang, J. A., Alaka, G. J., Hazelton, A., Gopalakrishnan, S. (2021) Coastal downwelling intensifies landfalling hurricanes, *Earth and Space Science Open Archive*, 1-21. *https://doi.org/10.1002/essoar.10508339.1*
16. Drake, L., 2012: Standardizing hurricane size descriptors for broadcast to the public. *40th Conference on Broadcast Meteorology, AMS https://ams.confex.com/ams/40BROADCAST/webprogram/Paper208126.html* (Accessed March 15, 2023).
17. Landsea, C. W., and J. L. Franklin, (2013) Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format. *Monthly Weather Review*, **141**, 3576–3592, *https://doi.org/10.1175/MWR-D-12-00254.1*.
18. NOAA's Atlantic Oceanographic and Meteorological Laboratory, U.S. Department of Commerce. Hurricanes Frequently Asked Questions, *https://www.aoml.noaa.gov/hrd-faq/#avg-forward-speed*
19. Demuth, J., DeMaria, M., and Knaff, J.A. (2006) Improvement of advanced microwave sounder unit tropical cyclone intensity and size estimation algorithms, *J. Appl. Meteorol. Climatol.* **45**, 1573-1581.
20. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane IKE, *https://www.nhc.noaa.gov/archive/2008/al09/al092008.public_b.048.shtml?* (Accessed Nov 2020)
21. National Weather Service, NOAA. Hurricane Ike-September 2008, *https://www.weather.gov/hgx/projects_ike08* (Accessed Nov 2020)

22. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane IVAN, *https://www.nhc.noaa.gov/archive/2004/pub/al092004.public_a.055.shtml?* (Accessed Nov 2020)
23. National Weather Service, NOAA. Powerful Hurricane Ivan Slams the Central Gulf Coast, *https://www.weather.gov/mob/ivan.* (Accessed Nov 2020)
24. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane KATRINA. *https://www.nhc.noaa.gov/archive/2005/pub/al122005.public_a.026.shtml* (Accessed Feb 2023)
25. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane RITA, *https://www.nhc.noaa.gov/archive/2005/pub/al182005.public_a.026.shtml?* (Accessed Nov 2020)
26. National Weather Service, NOAA. Hurricane Rita - September 2005, *https://www.weather.gov/hgx/projects_rita05* (Accessed Nov 2020)
27. Earth Observatory, NASA. Hurricane Dennis, *https://earthobservatory.nasa.gov/images/15093/hurricane-dennis* (Accessed Nov 2020)
28. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane DENNIS, *https://www.nhc.noaa.gov/archive/2005/pub/al042005.public_a.025.shtml?* (Accessed Nov 2020)
29. National Weather Service, NOAA. Hurricane Dennis - September 2005, *https://www.weather.gov/mob/dennis* (Accessed Nov 2020)
30. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane EARL, *https://www.nhc.noaa.gov/archive/1998/archive/pub/PAAL0598.010* (Accessed Nov 2020)
31. National Weather Service, NOAA. Hurricane Earl - September 3, 1998, *https://www.weather.gov/mob/earl* (Accessed Nov 2020)
32. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane GEORGES, *https://www.nhc.noaa.gov/archive/1998/archive/pub/PBAL0798.050* (Accessed Nov 2020)
33. National Weather Service, NOAA. Hurricane Georges - September 28, 1998, *https://www.weather.gov/mob/georges* (Accessed Nov 2020)
34. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane GUSTAV, *https://www.nhc.noaa.gov/archive/2008/al07/al072008.public_a.031.shtml?* (Accessed Nov 2020)
35. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane LILI, *https://www.nhc.noaa.gov/archive/2002/pub/al132002.public_b.048.html* (Accessed Nov 2020)
36. National Hurricane Center and Central Pacific Hurricane Center, NOAA. Hurricane CLAUDETTE, *https://www.nhc.noaa.gov/archive/2002/pub/al132002.public_b.048.html* (Accessed Nov 2020)
37. Hurricanes: Science and Society, NSF and University of Rhode Island. 1992 – Hurricane Andrew, *http://www.hurricanescience.org/history/storms/1990s/andrew/* (Accessed Nov 2020)
38. NOAA Shoreline Website, NOAA. NOAA Medium Resolution Shoreline, *https://shoreline.noaa.gov/data/datasheets/medres.html* (Accessed Oct 2020)
39. Gulf of Mexico Bathymetry Contours, USGS. *https://pubs.usgs.gov/of/2000/of00-352/data/bathyc/browse/bathyc.gif* (Accessed July 2022)
40. Hamilton, P. (2006) Deep-Current Variability near the Sigsbee Escarpment in the Gulf of Mexico, *Journal of Physical Oceanography* **37**, 708-726. *https://doi.org/10.1175/JPO2998.1*
41. *GraphPad Prism 9 Statistics Guide - Interpreting results: Mann-Whitney test. https://www.graphpad.com/guides/prism/latest/statistics/how_the_mann-whitney_test_works.htm (Accessed April 17, 2023).*
42. Wilks, D. (2011) Forecast Verification in *Statistical Methods in the Atmospheric Sciences* (Dmowska, R., Hartmann, D., Rossby, H. T., Ed.) 2nd Ed., 255-332, Elsevier.
43. Klotzbach, P. J., Bell, M. M., Bowen, S. G., Gibney, E. J., Knapp, K. R., Schreck, C. J. (2020) Surface Pressure a More Skillful Predictor of Normalized Hurricane Damage than Maximum Sustained Wind. *Bulletin of the American Meteorological Society* **101**, E830–E846, *https://doi.org/10.1175/BAMS-D-19-0062.1*
44. Ditchek, S. D., T. C. Nelson, M. Rosenmayer, and K. L. Corbosiero, (2017) The Relationship between Tropical Cyclones at Genesis and Their Maximum Attained Intensity. *Journal of Climate*, **30**, 4897–4913, *https://doi.org/10.1175/JCLI-D-16-0554.1*
45. Soleimany, A., E. Solgi, K. Ashrafi, R. Jafari, and R. Grubliauskas, (2022) Temporal and spatial distribution mapping of particulate matter in southwest of Iran using remote sensing, GIS, and statistical techniques. *Air Qual Atmos Health*, **15**, 1057–1078, *https://doi.org/10.1007/s11869-022-01179-y*
46. Bloemendaal, N., Muis, S., Haarsma, R. J., Verlaan, M., Irazoqui Apecechea, M., de Moel, H., Ward, P. J., Aerts, J. C. J. H. (2019) Global modeling of tropical cyclone storm surges using high-resolution forecasts. *Clim Dyn* **52**, 5031-5044. *https://doi.org/10.1007/s00382-018-4430-x*
47. Mori, N., Kato, M., Kim, S., Mase, H., Shibutani, Y., Takemi, T., Tsuboki, K., Yasuda, T. (2014) Local amplification of storm surge by Super Typhoon Haiyan in Leyte Gulf. *Geophysical Research Letters* **41**, 5106-5113. *https://doi.org/10.1002/2014GL060689*

ABOUT STUDENT AUTHOR

Kayleigh Addington graduated from Virginia Tech in May 2021 with a B.S. in Meteorology and a double major in Geography.

PRESS SUMMARY

Storm surge is the most dangerous component of landfalling tropical cyclones. The growing coastal population highlights the importance of research regarding the atmospheric and geographic factors influencing storm surge height in these storms. This study uses a sample of tropical cyclones making landfall in the Gulf of Mexico that are representative of the larger population to isolate the effect of coastline shape on maximum storm surge height. A measure of coastline concavity is developed using geometry and 50, 100, and 200 km radius circles around the point of maximum storm surge. Statistical tests are used to look for relationships between the maximum storm surge height and the shape of the coastline. While there is only a small difference in the maximum storm surge height on the two coastline types, results suggest that storm surges occurring on an inward-sloping (concave) coastline tend to be greater than storm surges occurring on an outward-sloping (convex) coastline. The number of factors influencing maximum storm surge height makes it difficult to investigate the influence of the angle of the coastline on storm surge height. The results are limited by the small number of storms and the numerous factors influencing maximum storm surge height.

Validation of a Computationally Efficient Model of the Mu-Opioid Receptor

Allison Barkdull[a], Lexin Chen[b], Akash Mathavan[c], Karina Martinez-Mayorga[d,e], & Coray M. Colina[b,e,f]*

[a]Department of Biomedical Engineering, University of Florida, Gainesville, FL
[b]Department of Chemistry, University of Florida, Gainesville, FL
[c]College of Medicine, University of Florida, Gainesville, FL
[d]Instituto de Química, Unidad Mérida, Universidad Nacional Autónoma de México, Ucú, Yucatán, México
[e] UAEY-IIMAS, Universidad Nacional Autónoma de México, Mérida, Yucatán, México
[f]Departments of Materials Science and Nuclear Engineering, University of Florida, Gainesville, FL

https://doi.org/10.33697/ajur.2023.085

Students: allisonbarkdull@ufl.edu, le.chen@ufl.edu, amathavan2496@ufl.edu
*Mentors: kmtzm@unam.mx, colina@chem.ufl.edu**

ABSTRACT
The mu-opioid receptor (MOR) is a transmembrane protein and the primary target for pain-modulating drugs. Opioid drugs come with detrimental side-effects such as physical dependence and addiction. However, recent studies show that understanding structural properties and dynamics of MOR may aid in the design of opioid drugs with reduced side effects. Molecular dynamics simulations allow researchers to study changes in protein conformation at an atomistic level. However, modeling systems including MOR embedded in a lipid bilayer can be computationally expensive. This study evaluates a modeling approach that uses harmonic restraints on the transmembrane regions of MOR to model the rigidity of the lipid bilayer without explicitly simulating lipid molecules, reducing the number of atoms in the simulation. The proposed model allows MOR to be simulated 49% faster than a simulation explicitly including the lipid bilayer. To assess the accuracy of the proposed model, simulations were performed of MOR in a lipid bilayer, the free MOR in water and MOR in water with harmonic restraints applied to all transmembrane residues using NAMD 3.0 alpha and the CHARMM36 force field. Dynamic properties of MOR were shown to be different in each system, with the free MOR having a higher root mean square deviation (RMSD) than MOR with an explicitly modeled lipid bilayer. The systems with harmonic restraint constants of 0.001 kcal/mol/Å^2 applied to the transmembrane residues had RMSD values comparable to those in an explicitly modeled lipid bilayer. This study demonstrates that using restraints on the transmembrane residues of MOR is a feasible way of modeling the ligand-free receptor with reduced computational costs. This model could allow the dynamics of MOR in a lipid bilayer environment to be studied more efficiently.

KEYWORDS
Molecular Dynamics; Atomistic Simulations; Computational Modeling; Mu-Opioid Receptor; G-Protein Coupled Receptor; Lipid Bilayer; Opioid; Transmembrane Protein

INTRODUCTION

Opioid receptors are a primary target for analgesic drugs. In 2020, there were 43.3 opiate prescriptions per 100 people in the United States.[1] However, opiates produce adverse side effects such as constipation, nausea, and respiratory depression, and long-term use of opioid drugs can lead to tolerance, physical dependence, and addiction.[2] The widespread abuse of opioid drugs is a major public health crisis which caused 50,178 deaths in the United States in 2019 alone[3]. Because of the limitations of existing opioid drugs, it is necessary to further explore how the opioid receptors can be modulated to minimize undesired effects.

Most pain-killing effects of opioid drugs are attributed to interactions with the mu-opioid receptor (MOR).[4] MOR, depicted in **Figure 1**, is a G protein-coupled receptor (GPCR) which is distributed through the nervous system and digestive tract. Like other GPCRs, MOR has seven transmembrane helices, an extracellular N-terminus and an intracellular C-terminus which can be phosphorylated upon agonist binding.[5] MOR has an orthosteric ligand-binding pocket, highlighted in purple in **Figure 1,** where both agonist and antagonist ligands bind. An agonist is a molecule which binds to a receptor causing a functional activation leading to a signaling cascade within the cell. In the case of MOR, an agonist binding is associated with activation of downstream signaling cascades that lead to a pain-killing effect. Conversely, an antagonist binds to the receptor blocking all other molecules from binding to the orthosteric site and locking the receptor in the inactive state. Upon activation of MOR by an agonist, there is a large outward movement of transmembrane helix 6 and a smaller inward movement of transmembrane helices 5 and 7.[6]

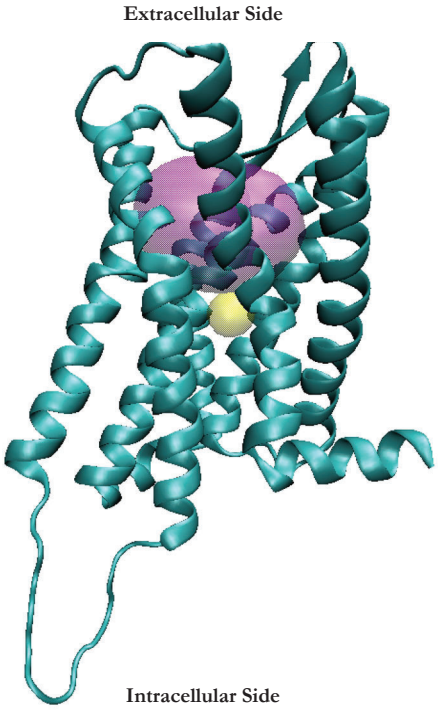


Figure 1. The mu-opioid receptor (PDB ID: 4DKL[22]) shown in blue with the agonist binding site represented in purple and the allosteric sodium binding site in yellow.

Evidence suggests that interactions with different agonists can cause significantly different conformational changes in MOR, leading to preferential activation of different downstream pathways.[7] This process is called functional selectivity or biased agonism.[7] MOR's analgesic effects are attributed to its interactions with heterotrimeric G-proteins[8] which act by inhibiting adenyl cyclase, modulating ion channels, and causing second messenger cascades.[9] However, after activation, MOR can be phosphorylated, leading to β-arrestin binding, which causes receptor desensitization and internalization which may be associated with negative side effects such as tolerance. It has been shown that different agonists cause different levels of G_i coupling and β-arrestin recruitment;[10] however, the exact conformational changes in MOR which lead to these different effects are not fully understood.[7] Finding ligands that can modulate pain with diminished unwanted downstream effects requires a better understanding of the structural changes of MOR when bound to different ligands.

MOR also has an allosteric binding site, highlighted in yellow on **Figure 1**, which plays an important role in the receptor conformation and downstream effects. This binding site is conserved among Class A GPCRs and may hold a sodium ion.[11] The presence of a sodium ion in the allosteric binding site helps stabilize MOR in its inactive state.[12] Site-directed mutagenesis studies replacing the residues which make up the binding site showed that there was significantly increased β-arrestin recruitment when sodium is not present[10] Additionally, computational studies suggest that herkinorin, an opioid which does not promote β-arrestin recruitment or receptor internalization may interact with the allosteric sodium binding site, modulating it.[13,14]

Molecular dynamics simulations are a powerful tool to study MOR dynamics and drug binding at an atomistic level. Molecular dynamics simulations have been used to gain understanding about agonist[15] and antagonist[16] binding properties, the role of the allosteric sodium ion[13,17] and functional selectivity.[18] In other studies, restraints on specific regions of integral membrane proteins have been used to represent features such as ion bridges[19] or to resolve the packing of transmembrane alpha helices in low-resolution crystal structures.[20] However, it can be challenging to study binding mechanisms of ligands to GPCRs such as MOR using all-atom molecular dynamics simulations due to the long time scales necessary to sample conformational states of GPCRs.[21]

This study seeks to evaluate a model of the ligand-free MOR which can be used to study its overall dynamic properties while eliminating the computational cost of explicitly modeling the lipid bilayer. The inactive MOR is simulated both with and without the allosteric sodium ion present in three conditions: the protein in an explicit lipid bilayer, the free protein surrounded by water, and the protein surrounded by water with a harmonic restraint applied to the backbone atoms of the transmembrane residues. Comparisons between the free protein and membrane-protein system allow for the evaluation of the impact of the membrane on protein dynamics. In order to evaluate the accuracy of the restrained protein model, comparisons are made with simulations of MOR in a membrane made of 1-palmitoyl-2-oleoyl phosphatidylcholine (POPC) and 10% cholesterol.

METHODS AND PROCEDURES

Figure 2 shows a summary of the simulation setups. The structure of the inactive mouse MOR was downloaded from the Protein Databank, with PDB ID: 4DKL.[22] In the original crystal structure, in the inactive state, the third intracellular loop was replaced with a T4 lysozyme. The sequence of the third loop is identical to that in the active mouse MOR structure (PDB ID: 5C1M[23]), so the structures were aligned and the residues from the active structure were added to the inactive structure, replacing the T4 lysozyme. While there are some differences between the mouse and human MOR, the amino acid sequences of the two proteins have a 96.2% similarity.[24, 25, 26] Then, the structure was minimized for 1000 steps to reach a more energetically favorable conformation. Each amino acid in the protein was modeled in the protonation state appropriate for a pH of 7.

The membrane-protein system was created by placing the protein in a membrane made of 1-palmitoyl-2-oleoyl phosphatidylcholine (POPC) and 10% cholesterol using the CHARMM-GUI Membrane Builder.[27] The system was buffered by 22.5 Å of cTIP3P[28] water on the top and bottom of the membrane (**Figure 2**, Panels A-B). The free protein system was solvated with cTIP3P[29] water molecules in a 111×111×111 $Å^3$ cubic box to prevent interactions with periodic images (**Figure 2**, C-D).

The restrained protein system was solvated in an orthorhombic box with a 16 Å cTIP3P water buffer on each side of the protein, and the transmembrane residues identified from the PDB entry were restrained during production with varying forces.[22] Five simulations were performed with restraint forces of 20 kcal/mol/Å^2, 5 kcal/mol/Å^2, 1 kcal/mol/Å^2, 0.01 kcal/mol/Å^2, and 0.001 kcal×mol^{-1}×Å^{-2} (**Figure 2**, Panel E-F) on the transmembrane atoms. The free protein system was created by solvating the protein in a 111Å × 111Å × 111Å cubic box of cTIP3P water with no restraints added during the production simulation.

In the restrained protein system, the receptor is held by the restraints in the center of the box and translates less than the free protein system, so a smaller water buffer can be used without the protein interacting with its own periodic images. This smaller water buffer contributes to a higher simulation speed.

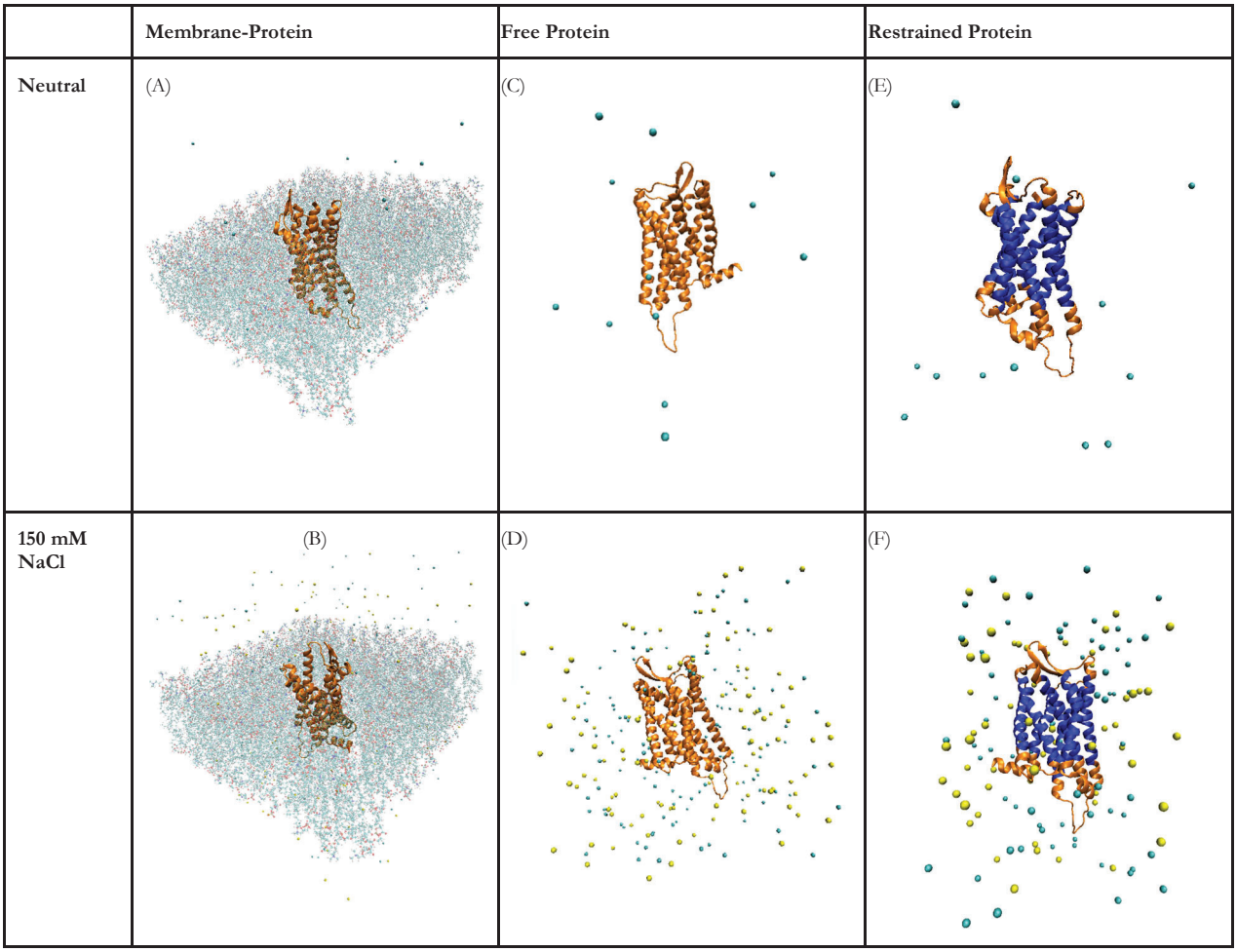


Figure 2. Summary of simulation setups. Although explicitly modeled, water molecules are not shown for clarity. Sodium ions are shown as yellow beads and chlorine ions are shown as blue beads (A) MOR (orange) is simulated in an explicit lipid bilayer with 22.5 Å of water on the top and bottom of the membrane. 13 chlorine ions are added to neutralize the system. (B) MOR (orange) is simulated in an explicit lipid bilayer with 22.5 Å of water on the top and bottom of the membrane and 150 mM sodium chloride. (C) The free MOR is simulated in a 111 Å^3 water box neutralized with 13 chlorine ions. (D) The free MOR is simulated in a 111 Å^3 water box with 150 mM sodium chloride. (E) MOR is simulated within an orthorhombic box with a 16 Å water buffer and transmembrane residues (blue) restrained. 13 chlorine ions are added to neutralize the system. (F) MOR is simulated within an orthorhombic box with a 16Å water buffer and 150 mM sodium chloride. Transmembrane residues (blue) are restrained.

Each of the three systems were solvated in water with 13 chlorine counterions to neutralize the system and in a 150 mM NaCl solution. Systems containing 13 chlorine counterions and no NaCl will be referred to as "neutral pH" and systems with 13 chlorine counterions and 150 mM NaCl will be referred to as "150 mM NaCl". In all the systems solvated with NaCl, the allosteric sodium ion was manually moved to its binding site at the midpoint of D114 and S154. The box dimensions for each simulation are summarized in **Table 1**.

Atomistic molecular dynamics simulations were performed using the NAMD 3.0 alpha software package.[30] The CHARMM36m

force field[31] was used to model the protein and the CHARMM C36 lipid force field[32] was used to model the membrane. Throughout all of the simulations, the SETTLE algorithm[33] was used to constrain bonds that include a hydrogen atom and a 2 fs time step was employed. In all simulations, a switching distance of 10 Å and cutoff of 12 Å was used for all nonbonded interactions. Long-range electrostatics beyond the nonbonded cutoff of 12 Å were calculated using the particle mesh Ewald (PME) summation. All simulations were performed at 303.15 K. Each simulation was performed using one NVIDIA A100 Tensor Core GPU on Hipergator 3.0.

	Membrane-Protein	**Free Protein**	**Restrained Protein**
Box Dimensions	120 Å×120 Å×125 Å	111 Å×111 Å×111 Å	81 Å×73 Å×111 Å

Table 1. Box dimensions for each system setup.

The membrane-protein systems were minimized for 10,000 steps using the conjugate gradient algorithm with harmonic and dihedral restraints on the protein and lipid and a planar restraint on the membrane. The system was gradually heated from 0 to 303.15 K in an NVT (constant number of atoms, volume, and temperature) ensemble with harmonic restraints on protein and lipids. Minimization was followed by six steps of equilibration with NPAT (constant number of atoms, pressure, lipid bilayer area, and temperature) dynamics. During equilibration, harmonic restraints were applied to the ions and protein backbone and planar restraints were used to prevent water from entering the membrane and to hold the lipid head groups in the plane of the Z-axis. Restraint force constants used during minimization and equilibration are summarized in **Table 2**. Production simulations were performed in a NPAT ensemble with pressure only changing in the Z direction using the Langevin thermostat with a $1ps^{-1}$ collision frequency and Nosé-Hoover Langevin barostat with a timestep of 2 fs.

Step	**Ensemble**	**Time (ps)**	**Backbone Heavy Atoms Restraint***	**Sidechain Heavy Atoms Restraint***	**Lipid Position Restraint***	**Lipid Dihedral Restraint***
Minimization	NVT	20	10	5.0	5	500
1	NPT	250	10	5.0	5	500
2	NPAT	250	5.0	2.5	5	200
3	NPAT	500	2.5	1.25	2	100
4	NPAT	500	1.0	0.5	1	100
5	NPAT	500	0.5	0.25	0.2	50
6	NPAT	500	0.1	0.05	0	0

Table 2. Summary of minimization and equilibration setup for the membrane-protein system. *All units for restraint constants are in kcal/mol/Å^2

The free protein systems were minimized for 10,000 steps using the conjugate gradient algorithm with a 20 kcal/mol/Å^2 harmonic force restraining the protein backbone and the system was heated from 0 K to 303.15 K in an NVT ensemble. This was followed by four stages of equilibration, each lasting 25,000 steps, with the harmonic restraint constant being gradually decreased from 20 kcal/mol/Å^2 to 10 kcal/mol/Å^2 to 5 kcal/mol/Å^2 to 2.5 kcal/mol/Å^2 in an NPT (constant number of atoms, pressure, and temperature) ensemble with a 2 fs time step. The production data was generated using an NPT ensemble at 303.15 K and 1 atm using the Langevin thermostat with a 1 ps^{-1} collision frequency and Nosé-Hoover Langevin barostat with a 2 fs time step and coordinates and system properties output every 10 ps.

The restrained protein simulations were minimized and heated using the same methods as the free protein systems. This was followed by four stages of equilibration, each lasting 25,000 steps with the protein restraint being gradually decreased from 20 kcal/mol/Å^2 to the desired restraint in an NPT ensemble with a 2 fs time step. The production simulations were performed with a harmonic restraining force applied to all atoms in transmembrane residues of MOR from the protein with PDB ID: 4DKL.[34] The harmonic restraining forces reduce mobility by adding an extra energetic penalty defined by **Equation 1**.

$$U(x) = k(x - x_{ref})^2$$ **Equation 1.**

Where x is the position of an atom at a given time, x_{ref} is the position of that atom at the first frame of minimization and k is the force constant. As mentioned above, the force constants used in this study were 20 kcal/mol/Å^2, 5 kcal/mol/Å^2, 1 kcal/mol/Å^2, 0.01 kcal/mol/Å^2 and 0.001 kcal/mol/Å^2, and were applied during production simulation to all atoms in the transmembrane region of MOR. Other than the restraints, production data was generated using the same methods as the free protein systems.

For all systems, results were generated from simulations of 1000 ns with the first 200 ns being considered as equilibration and not included in calculations of average values. System properties were output every 50000 steps. The root mean square deviation (RMSD) and root mean squared fluctuation (RMSF) were calculated using cpptraj.[35] In this study, the RMSD is a measure of the difference between the protein coordinates at a reference state, and the protein coordinates at a given timestep. The RMSD is calculated by

$$RMSD = \sqrt{\sum_{i=0}^{N} \frac{m_i(X_i - Y_i)^2}{M}} \qquad \textbf{Equation 2.}$$

Where m_i is the mass of a given atom, X_i is the position of a given atom, Y_i is the reference position of a given atom and M is the total mass of the system. In this work, RMSDs are calculated for the backbone atoms (nitrogen, carbon, and alpha carbon) and the first frame after minimization and heating is used as the reference structure.

The RMSF is the time average of the RMSD and is calculated for each individual residue. The RMSF for an atom X_i is calculated by

$$RMSF = \sqrt{\langle (X_i - \langle X_i \rangle)^2 \rangle} \qquad \textbf{Equation 3.}$$

Where X_i is the coordinates of as particle and $\langle X_i \rangle$ is the average position of that particle. In this work RMSF is calculated for backbone atoms using the first frame after minimization and heating as the reference structure.

Additionally, the distributions of the χ_1 and χ_2 angles of N332 were calculated using cpptraj.[35] The χ_1 angle is the torsional angle including the nitrogen, α-carbon, β-carbon, and γ-carbon of N332 and the χ_2 angle is the torsional angle involving the α-carbon, β-carbon, and γ-carbon, and δ-oxygen of N332. The confirmation of these torsional angles has been shown to be related to the downstream signaling of the receptor.[13]

All error values are reported as a 95% confidence interval which is calculated as $\pm$ 1.96σ, where σ is the standard deviation.

RESULTS AND DISCUSSION

The RMSD of the systems studied in this work are shown in **Figure 3**. To note, the RMSD of the free MOR system (**Figure 3**, panel B) is higher than the RMSD of MOR in its explicitly modeled lipid bilayer (**Figure 3**, panel A). The average RMSD of the free protein is 3.9 ± 0.5 Å and 3.4 ± 0.6 Å for the neutral system and system with 150 mM NaCl respectively, compared to the membrane-protein system which has an average RMSD of 3.1 ± 0.6 Å and 3.2 ± 0.3 Å for the neutral system and system with 150 mM NaCl respectively (**Table 3**). Thus, as expected, the protein moves more freely in the absence of the lipid bilayer. This is because the amino acids in the transmembrane regions of the lipid-free MOR are not structurally stabilized by the viscous lipid bilayer, allowing more freedom of motion. However, adding restraints to the transmembrane region lowers the overall RMSD (**Figure 3**, panel C) for all of the restraint force constants tested. The RMSD of the protein with the 0.001 kcal/mol/Å^2 restraint force constant is closest to the RMSD of the protein with its explicitly modeled lipid bilayer (**Table 3**). Systems with higher restraint force constants (0.01–20 kcal/mol/Å^2) all have lower average RMSD values than the membrane-protein system. Additionally, the RMSD of systems simulated with 150 mM of NaCl tend to have a lower RMSD than systems in neutral conditions without the allosteric sodium ion; this is in agreement with previous studies that show that the allosteric sodium ion stabilizes MOR into its inactive state.[11, 12]

System	RMSD of System with Neutral pH (Å)	RMSD of System with 150 mM NaCl (Å)
Free Protein	3.9 ± 0.5	3.4 ± 0.6
Membrane-Protein	3.1 ± 0.6	3.2 ± 0.3
MOR with Restraints on Transmembrane Atoms		
20 kcal/mol/Å²	1.9 ± 0.5	2.2 ± 0.4
5 kcal/mol/Å²	2.3 ± 0.6	2.1 ± 0.5
1 kcal/mol/Å²	2.3 ± 0.6	2.2 ± 0.5
0.01 kcal/mol/Å²	2.4 ± 0.6	2.9 ± 1.0
0.001 kcal/mol/Å²	3.1 ± 0.5	3.0 ± 0.4

Table 3. Summary of average RMSDs for each system. The average RMSD is calculated from the last 800 ns of simulation and error is reported as the 95% confidence interval.

To better understand how the dynamic properties of MOR differ in the three models, the RMSD was calculated only for the transmembrane residues, *i.e.*, residues 67 to 91, 105 to 129, 141 to 163, 184 to 205, 229 to 253, 278 to 304, and 313 to 336 and represented in **Figure 4** and **Table 4**. This shows that the RMSD of the transmembrane residues of the free MOR are 1–1.3 Å higher than in MOR in an explicitly modeled lipid bilayer. The average transmembrane RMSDs in the free protein system are 3.2 ± 0.4 Å and 2.3 ± 0.6 Å for the neutral system and the system with 150 mM NaCl respectively, compared to 1.9 ± 0.3 Å and 1.3 ± 0.2 Å in the system with the explicitly modeled lipid bilayer (**Table 4**).

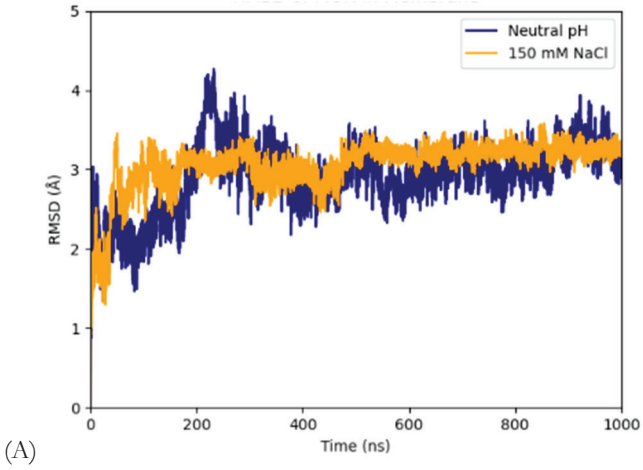
Addition of restraints successfully lowers the RMSD of the transmembrane residues as shown in **Figure 4**, panel C. However, the systems with 1 kcal/mol/Å², 5 kcal/mol/Å², and 20 kcal/mol/Å² restraint constants have drastically lower RMSDs in the transmembrane region (**Table 4**) and do not accurately model the dynamics of MOR in an explicitly modeled lipid bilayer. The protein with the 0.01 kcal/mol/Å² restraint constant is also notably lower than that of MOR in its explicitly modeled lipid bilayer. However, the protein with the 0.001 kcal/mol/Å² restraint constant has a transmembrane RMSD that differs only by 0.3 Å from the membrane-protein system for both the neutral system and the system with 150 mM NaCl. Throughout the simulations, the transmembrane RMSD of the systems with the allosteric sodium ion tend to be lower due to its stabilizing effect on the inactive MOR.[11, 12]

System	RMSD of Transmembrane Atoms of System with Neutral pH (Å)	RMSD of Transmembrane Atoms of System with 150 mM NaCl (Å)
Free Protein	3.2 ± 0.4	2.3 ± 0.6
Membrane-Protein	1.9 ± 0.3	1.3 ± 0.2
MOR with Restraints on Transmembrane Atoms		
20 kcal/mol/Å²	0.12 ± 0.01	0.13 ± 0.01
5 kcal/mol/Å²	0.27 ± 0.02	0.18 ± 0.02
1 kcal/mol/Å²	0.27 ± 0.02	0.28 ± 0.02
0.01 kcal/mol/Å²	0.92 ± 0.11	0.87 ± 0.15
0.001 kcal/mol/Å²	1.6 ± 0.3	1.6 ± 0.3

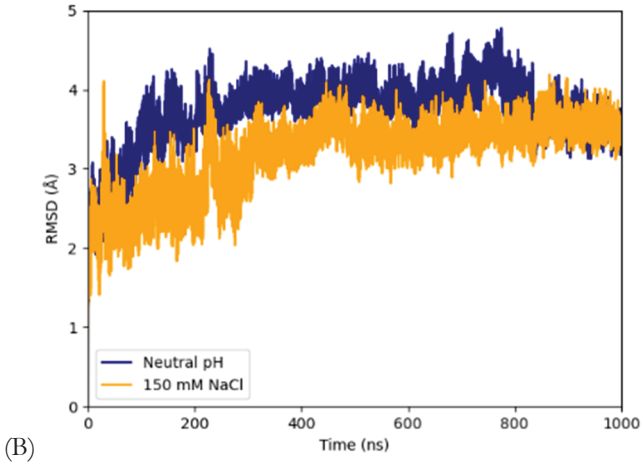
Table 4. Summary of average RMSDs of transmembrane residues for each system. The average RMSD is calculated from the last 800ns of simulation and error is reported as the 95% confidence interval.

The RMSF of each system is analyzed to showcase the impact of each system setup in different regions of the protein and shown in **Figure 5**. Across all residues, the free MOR tends to have a higher RMSF than the membrane-protein system.

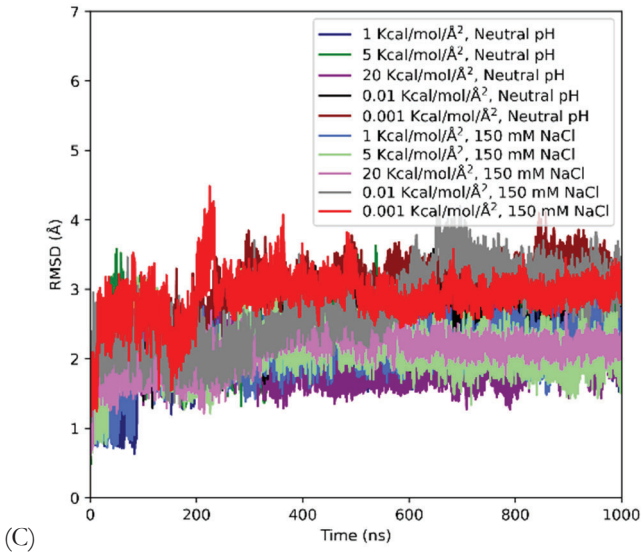
Of the systems with transmembrane restraints, the system with the 0.001 kcal/mol/Å² restraint force constant has an RMSF that most closely models the membrane-protein system in the transmembrane region. However, the protein with the 0.001 kcal/mol/Å² restraint force constant tends to have a higher RMSF in the intra and extracellular regions than the membrane protein system. Particularly, the RMSF of residues 254-277, which are highlighted in purple on **Figure 6**, and residues 305-312, which are highlighted in green on **Figure 6**, are higher in the restrained protein system than the membrane-protein system. It is possible that the higher RMSF in the system with the 0.001 kcal/mol/Å² restraint force constant could be due to differences caused by the lack of an explicitly modeled lipid bilayer. However, both of these regions are mostly in a random coil structure and therefore more flexible than other regions of the protein, providing a possible explanation for greater differences between simulations.



(A)



(B)



(C)

Figure 3. (A) The RMSD of MOR in an explicitly modeled lipid bilayer (B) The RMSD of the free MOR (C) The RMSD of MOR with restraints

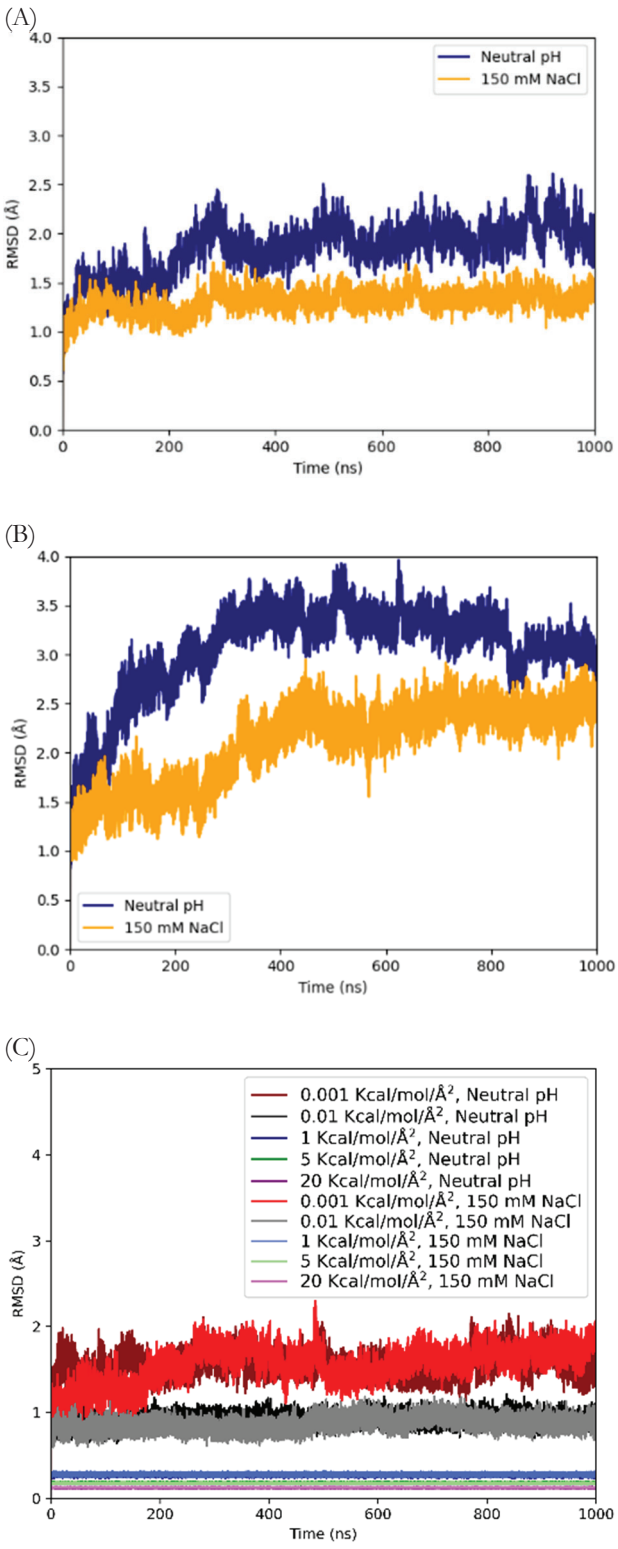
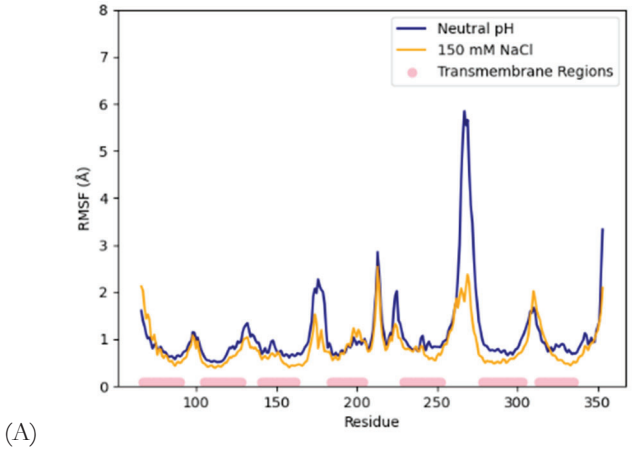
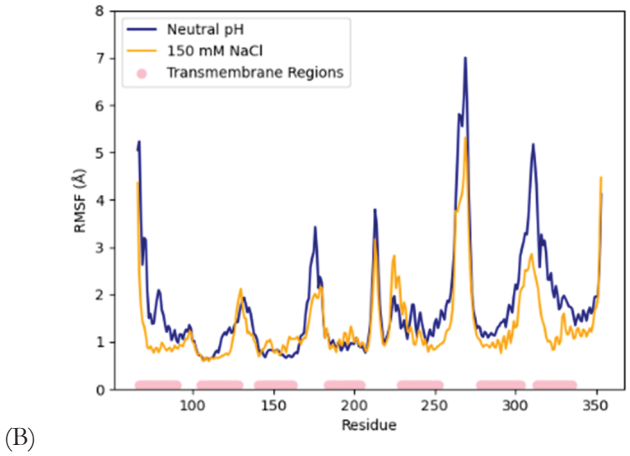


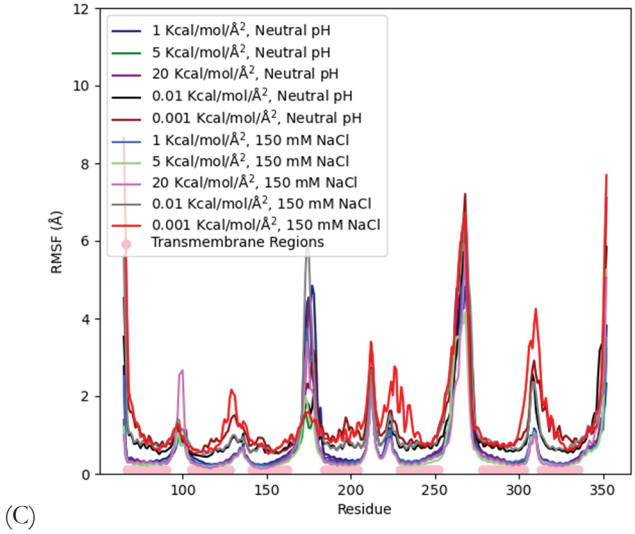
Figure 4. (A) The RMSD of transmembrane residues of the membrane-protein system (B) The RMSD of transmembrane residues of the free MOR (C) The RMSD of transmembrane residues of MOR with restraints.



(A)



(B)



(C)

Figure 5. All RMSFs are calculated from the last 800 ns of simulation data. The transmembrane residues are highlighted in pink. (A)The RMSF of MOR in the membrane-protein system (B) The RMSF of the free MOR (C) The RMSF of the restrained MOR

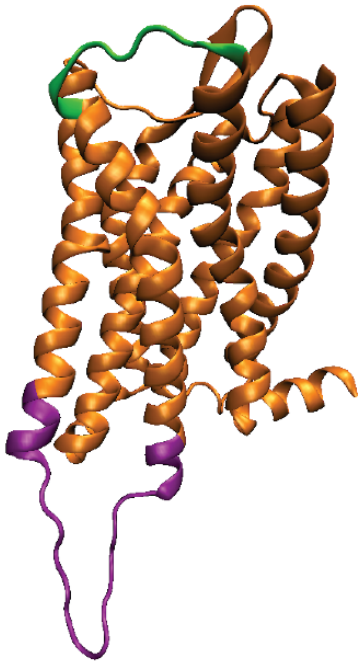


Figure 6. MOR with regions of interest highlighted. Residues 254-277 are highlighted in purple and residues 305-312 are highlighted in green.

The impact of each method of modeling the lipid bilayer on the NPxxY motif was analyzed by calculating distributions of side-chain dihedral angles χ_1 and χ_2 of N332, shown in **Figure 7**. The conformation of the side-chain dihedral angles of N332 have been shown to be related with the downstream signaling of the receptor.[13]

In the system simulated with an explicitly modeled lipid bilayer, the χ_1 dihedral distribution has peaks near -75 and -180 degrees in the neutral system, and a peak at -75 in the system with 150 mM NaCl. In both the free MOR system and system with restraints on the transmembrane residues a peak at -80 degrees is observed in both the neutral system and the system with 150 mM NaCl. In the neutral free MOR system and neutral system with a 0.001kcal/mol/Å^2 restraint force constant, a peak is noticeable at -180 degrees, but the peak is smaller than the one observed in the system with the explicitly modeled lipid bilayer.

The χ_2 dihedral distribution of the system simulated with the explicitly modeled lipid bilayer has peaks of -20, and -115 degrees for the neutral system and a peak at -20 degrees for the system with 150 mM NaCl. Both the free MOR system and system with restraints on the transmembrane residues have peaks at -20 degrees for both the neutral and systems with -150 mM NaCl. A small peak at -115 degrees is observed in both the neutral free MOR system and the neutral system with a 0.001kcal/mol/Å^2 restraint force constant.

Ultimately, the systems with a 0.001kcal/mol/Å^2 restraint force constant on the transmembrane residues are able to capture the peaks in the distribution of the χ_1 and χ_2 torsion angles of N332. Because the χ_1 and χ_2 torsion angles of N332 are associated with the downstream signaling of MOR, this analysis should be used in future simulations containing the receptor a ligand bound state to validate the system with transmembrane restraints' ability to capture differences associated with downstream signaling.[13]

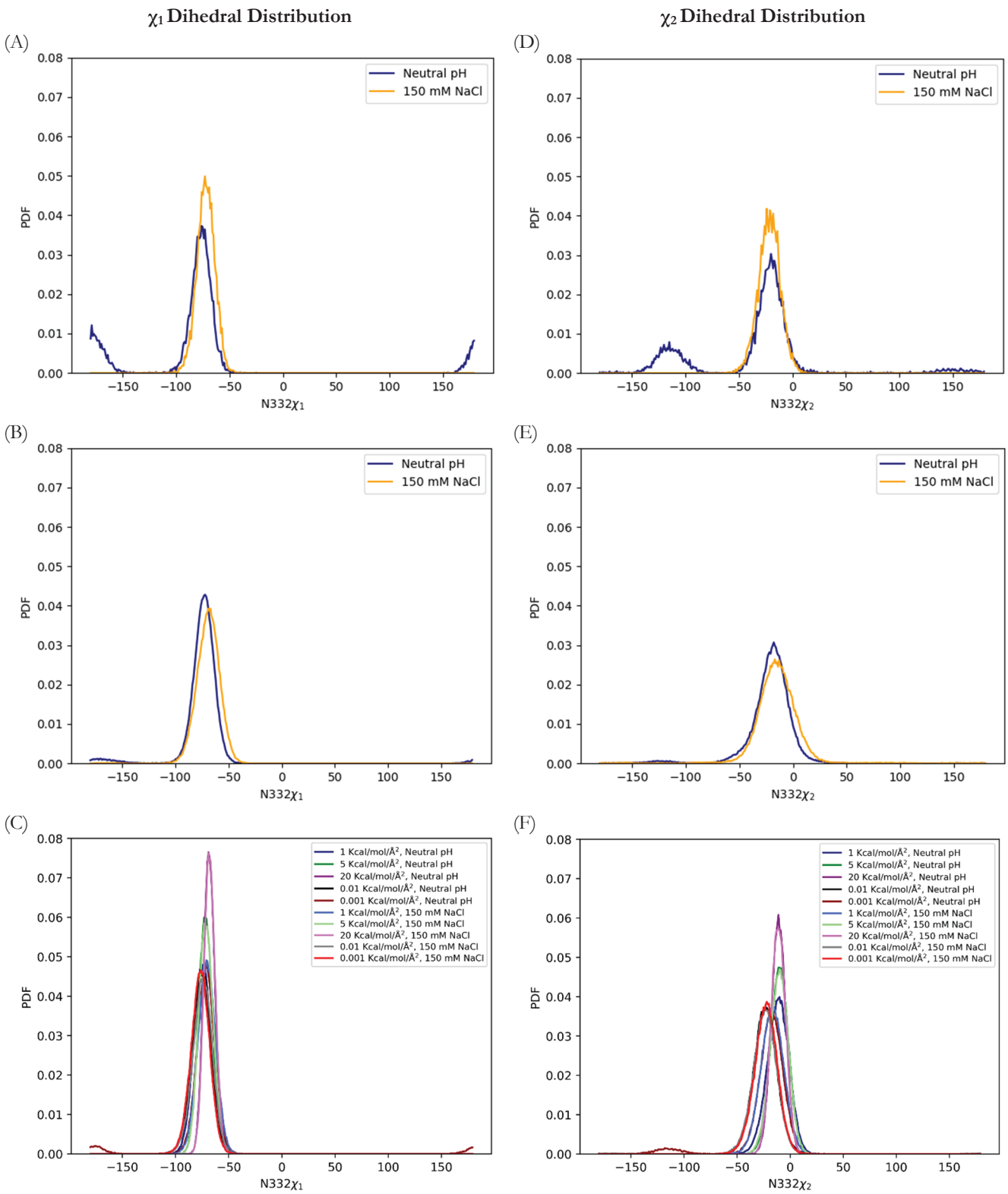


Figure 7. Dihedral distributions of the χ_1 (A-C) and χ_2 (D-F) angle of N332 are calculated from the last 800 ns of simulation data. (A) The dihedral distributions of the χ_1 angle of N332 of MOR in the membrane-protein system (B) The dihedral distributions of the χ_1 angle of N332 of the free MOR (C) The dihedral distributions of the χ_1 angle of N332 of the restrained MOR (D) The dihedral distributions of the χ_2 angle of N332 of MOR in the membrane-protein system (E) The dihedral distributions of the χ_2 angle of N332 of the free MOR (F) The dihedral distributions of the χ_2 angle of N332 of the restrained MOR

The values reported in this work for RMSD and transmembrane RMSD agree with values reported in literature. Ricarte *et al.* reports a total RMSD for MOR bound to morphine of 2.9-3.2 Å based on three simulations of 2-3 μs.[36] This coincides with the values reported in this work of 3.1-3.2 Å for the membrane-protein system and 3.0-3.1 Å for the system with 0.001 kcal/mol/Å^2 restraint constant. Cong *et al.* reports a transmembrane RMSD of 1.7, 1.3 and 1.2 Å for MOR bound to morphine, hydromorphone and β-funaltrexamine respectively[37], which agrees with the transmembrane RMSDs of 1.3-1.9 Å for the membrane-protein system and 1.6 Å for the system with 0.001 kcal/mol/Å^2 restraint force constant. One difference between the simulations in this work and the literature values summarized here is the presence of morphine bound to MOR. However, other simulations show that morphine does not have a dramatic impact on the RMSD of MOR.[38] Another difference between the results summarized here is that MOR was simulated in a membrane of 100% POPC whereas in this work MOR was simulated in a membrane of 90% POPC and 10% cholesterol. Short 200 ns simulations were performed using the same simulation setup outlined in this paper but with a membrane of 100% POPC to evaluate the role of bilayer composition on receptor dynamics. These simulations showed that the POPC membrane did not cause a dramatic difference in RMSD within the timescale, suggesting that the restrained-protein model may be appropriate to model bilayers with a range of POPC/cholesterol ratios (Supplemental Information, **Figure S1**).

Despite accurately modeling overall structural properties of MOR in its explicitly modeled lipid bilayer, there are some limitations to the restrained protein model. It has been shown that cholesterol can bind to residues on TM6 and TM7 of MOR affecting the likelihood of β-arrestin recruitment.[39] Without explicitly modeling the lipid bilayer, the impact of cholesterol binding cannot be modeled using the system proposed in this study. Additionally, experimental work has shown that different ligands bound to MOR can affect the protein's mobility perpendicular to the lipid bilayer.[40] Because the residues with restraints applied remain constant throughout the simulation, this is a feature of MOR-ligand interactions that could not be studied using the restrained protein model.

Since the restrained protein system with a 0.001 kcal/mol/Å^2 restraint force constant accurately models the dynamic properties of MOR in its explicitly modeled lipid bilayer, it is important to assess the relative computational efficiency of these systems. The restrained protein system is able to generate 49% faster than the system with the explicitly modeled lipid bilayer (**Figure 8**). The restrained protein system has the fewest number of atoms because it does not contain an explicitly modeled lipid bilayer, and the smaller size of the system contributes to increased simulation speed. This shows that using a backbone restraint on transmembrane residues can speed up data generation and increase the length of simulations of MOR.

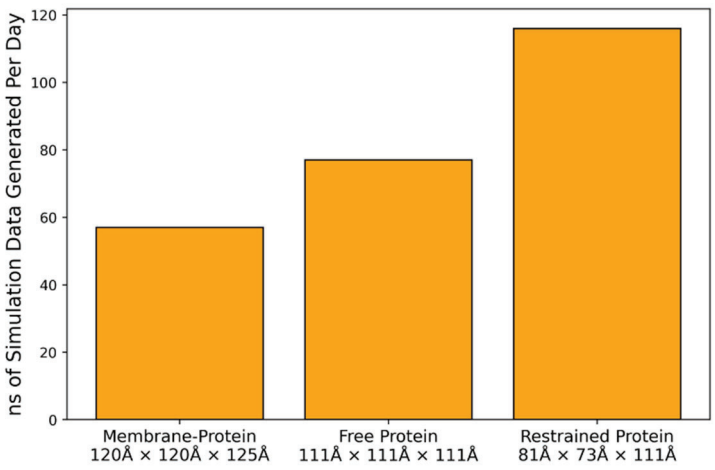


Figure 8. Nanoseconds of simulation data generated per day for each system.

CONCLUSIONS

This work shows that the free MOR in water has a higher RMSD than MOR in its explicitly modeled lipid bilayer, particularly in the transmembrane regions. Comparisons between MOR in its explicitly modeled lipid bilayer and MOR with varying restraints applied to its transmembrane residues show that using a restraint force constant of 0.001 kcal/mol/Å^2 on the transmembrane residues of MOR can model the overall dynamic properties of the membrane-protein system. The average RMSD values of MOR in its explicitly modeled lipid bilayer are reproduced within 95% confidence intervals by the corresponding systems simulated with a restraint force constant of 0.001 kcal/mol/Å^2 on the transmembrane residues.

The RMSF does differ between the restraint model and the membrane-protein model in the intra and extracellular regions, particularly residues 254-277 and residues 305-312, which is likely due to the highly flexible nature of this region. Additionally, analysis of the distribution of the χ_1 and χ_2 angles of N332 showed that simulations of the MOR with a 0.001kcal/mol/Å^2 restraint force constant on the transmembrane residues are able to capture the distributions seen in the system with an explicitly modeled lipid bilayer. Importantly, while using the restrained model data is generated twice as quickly as the explicit membrane model. With further validation of simulations using membranes on the transmembrane residues to simulate MOR bound to ligands, this method may be an effective approach to model overall structural properties of MOR, allowing for a faster screening of new opioid drug molecules.

In order to explore the applications of the restrained protein model in studies involving new opioid molecules and to assess its ability to model drug-induced protein dynamics, an opioid agonist such as morphine can be docked to the crystal structure of MOR using e.g., AutoDock Vina[41, 42] to create a starting structure that approximates the ligand's binding pose. The 3-dimensional structures of drug molecules can be taken from the Zinc Database. Parameters should be developed for the drug molecules using software such as the ParamChem[43, 44] tool for CGENFF 2.5.[45] The accuracy of the parameters for the drug molecules should also be assessed based on a comparison of the computational binding free energy with experimental data. If necessary, parameters can be further optimized using the ForceField Toolkit in VMD.[46, 47] Once the parameters have been validated, they can then be used to perform molecular dynamics simulations of at least 500 ns of the drug-receptor complex both in an explicitly modeled lipid bilayer and with a restraint force constant of 0.001 kcal/mol/Å^2 applied to the transmembrane residues. The results of the membrane-protein simulation and the restrained protein simulation should be compared to assess if the restrained model is able to represent protein dynamics associated with ligand binding.

ACKNOWLEDGEMENTS

Part of this research was financially supported by the University Scholars Program awarded to Allison Barkdull. Additionally, the authors thank the University of Florida Department of Chemistry and The J. Crayton Pruitt Family Department of Biomedical Engineering. Computational resources utilized were managed by the University of Florida (HiPerGator 3.0).

REFERENCES

1. CDC Injury Center, U.S. Opioid Dispensing Rate Maps | Drug Overdose | *https://www.cdc.gov/drugoverdose/rxrate-maps/index.html* (Accessed April 2022)

2. Minami, M. and Satoh, M. (1995) Molecular Biology of the Opioid Receptors: Structures, Functions and Distributions, *Neurosci. Res.* 23 (2), 121–145. *https://doi.org/10.1016/0168-0102(95)00933-k*

3. Ciccarone, D. (2021) The Rise of Illicit Fentanyls, Stimulants and the Fourth Wave of the Opioid Overdose Crisis, *Curr. Opin. Psychiatry.* 34 (4), 344–350. *https://doi.org/10.1097/yco.0000000000000717*

4. Matthes, H. W. D., Maldonado, R., Simonin, F., Valverde, O., Slowe, S., Kitchen, I., Befort, K., Dierich, A., Le Meur, M., Dollé, P., Tzavara, E., Hanoune, J., Roques, B. P., and Kieffer, B. L. (1996) Loss of Morphine-Induced Analgesia, Reward Effect and Withdrawal Symptoms in Mice Lacking the μ-Opioid-Receptor Gene, *Nature* 383 (6603), 819–823. *https://doi.org/10.1038/383819a0*

5. Minami, M., and Satoh, M. (1995) Molecular Biology of the Opioid Receptors: Structures, Functions and Distributions, *Neurosci. Res. 23* (2), 121–145. *https://doi.org/10.1016/0168-0102(95)00933-k*

6. Weis, W. I., and Kobilka, B. K. (2008) Structural Insights into G-Protein-Coupled Receptor Activation, *Curr. Opin. Struct. Biol.* 18 (6), 734–740. *https://doi.org/10.1016/j.sbi.2008.09.010*

7. Raehal, K. M., Schmid, C. L., Groer, C. E., and Bohn, L. M. (2011) Functional Selectivity at the μ-Opioid Receptor: Implications for Understanding Opioid Analgesia and Tolerance, *Pharmacol. Rev.* 63 (4), 1001–1019. *https://doi.org/10.1124/pr.111.004598*

8. Raffa, R. B., Martinez, R. P., and Connelly. (1994) C. D. G-Protein Antisense Oligodeoxyribonucleotides and μ-Opioid Supraspinal Antinociception, *Eur. J. Pharmacol.* 258(1-2), R5–R7. *https://doi.org/10.1016/0014-2999(94)90073-6*

9. Jordan, B., and Devi, L. A. (1998) Molecular Mechanisms of Opioid Receptor Signal Transduction. *Br. J. Anaesth.* 81 (1), 12–19. *https://doi.org/10.1093/bja/81.1.12*

10. Fenalti, G., Giguere, P. M., Katritch, V., Huang, X.-P., Thompson, A. A., Cherezov, V., Roth, B. L., and Stevens, R. C. (2014) Molecular Control of δ-Opioid Receptor Signalling, *Nature* 506 (7487), 191–196. *https://doi.org/10.1038/nature12944*

11. Katritch, V., Fenalti, G., Abola, E. E., Roth, B. L., Cherezov, V., and Stevens, R. C. (2014) Allosteric Sodium in Class a GPCR Signaling, *Trends Biochem. Sci.* 39 (5), 233–244. *https://doi.org/10.1016/j.tibs.2014.03.002*

12. Selley, D. E., Cao, C.-C., Liu, Q., and Childers, S. R. (2000) Effects of Sodium on Agonist Efficacy for G-Protein Activation in μ-Opioid Receptor-Transfected CHO Cells and Rat Thalamus. *Br. J. Pharmacol.* 130 (5), 987–996. *https://doi.org/10.1038/sj.bjp.0703382*

13. Marmolejo-Valencia, A. F., Madariaga-Mazón, A., and Martinez-Mayorga, K. (2021) Bias-Inducing Allosteric Binding Site in Mu-Opioid Receptor Signaling, *SN Appl. Sci.* 3 (5). https://doi.org/10.1007/s42452-021-04505-8.

14. Marmolejo-Valencia, A. F., and Martínez-Mayorga, K. (2017) Allosteric Modulation Model of the Mu Opioid Receptor by Herkinorin, a Potent Not Alkaloidal Agonist, *J. Comput. Aided Mol. Des.* 31 (5), 467–482. *https://doi.org/10.1007/s10822-017-0016-7*

15. Huang, W., Manglik, A., Venkatakrishnan, A. J., Laeremans, T., Feinberg, E. N., Sanborn, A. L., Kato, H. E., Livingston, K. E., Thorsen, T. S., Kling, R. C., Granier, S., Gmeiner, P., Husbands, S. M., Traynor, J. R., Weis, W. I., Steyaert, J., Dror, R. O., and Kobilka, B. K. (2015) Structural Insights into μ-Opioid Receptor Activation, *Nature* 524 (7565), 315–321. *https://doi.org/10.1038/nature14886*

16. Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K., and Granier, S. (2012) Crystal Structure of the μ-Opioid Receptor Bound to a Morphinan Antagonist, *Nature* 485 (7398), 321–326. *https://doi.org/10.1038/nature10954*

17. Shang, Y., LeRouzic, V., Schneider, S., Bisignano, P., Pasternak, G. W., and Filizola, M. (2014) Mechanistic Insights into the Allosteric Modulation of Opioid Receptors by Sodium Ions, *Biochemistry* 53 (31), 5140–5149. *https://doi.org/10.1021/bi5006915*

18. Hernández-Alvarado, R. B., Madariaga-Mazón, A., Cosme-Vela, F., Marmolejo-Valencia, A. F., Nefzi, A., and Martinez-Mayorga, K. (2021) Encoding Mu-Opioid Receptor Biased Agonism with Interaction Fingerprints, *J. Comput. Aided Mol. Des.* 35 (11), 1081–1093. *https://doi.org/10.1007/s10822-021-00422-5*

20. Shen, R., Han, W., Fiorin, G., Islam, S. M., Schulten, K., & Roux, B. (2015). Structural Refinement of Proteins by Restrained Molecular Dynamics Simulations with Non-interacting Molecular Fragments. *PLOS Comput. Biol. 11*(10), e1004368. https://doi.org/10.1371/journal.pcbi.1004368

21. Sansom, M. S. P., Sankararamakrishnan, R., and Kerr, I. D. (1995). Modelling membrane proteins using structural restraints. *Nature Structural Biology*, *2*(8), 624–631. https://doi.org/10.1038/nsb0895-624

21. Niesen, M. J. M., Bhattacharya, S., and Vaidehi, N. (2011) The Role of Conformational Ensembles in Ligand Recognition in G-Protein Coupled Receptors, *J. Am. Chem. Soc.*133 (33), 13197–13204. *https://doi.org/10.1021/ja205313h*

22. Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K., and Granier, S. (2012) Crystal Structure of the Mu-Opioid Receptor Bound to a Morphinan Antagonist. *https://doi.org/10.2210/pdb4dkl/pdb*

23. Huang, W. J., Manglik, A., Venkatakrishnan, A. J., Laeremans, T., Feinberg, E. N., Sanborn, A. L., Kato, H. E., Livingston, K. E., Thorsen, T. S., Kling, R., Granier, S., Gmeiner, P., Husbands, S. M., Traynor, J. R., Weis, W. I., Steyaert, J., Dror, R. O., and Kobilka, B. K. (2015) Crystal Structure of Active Mu-Opioid Receptor Bound to the Agonist BU72. *https://doi.org/10.2210/pdb5c1m/pdb*

24. Kaufman, Daniel L., et al. (1995) Characterization of the Murine μ Opioid Receptor Gene, *Journal of Biological Chemistry.* 270 (26), 15877–15883. https://doi.org/10.1074/jbc.270.26.15877.

25. Mestek, A, et al. (1995) The Human Mu Opioid Receptor: Modulation of Functional Desensitization by Calcium/Calmodulin-Dependent Protein Kinase and Protein Kinase C, *The Journal of Neuroscience.* 15 (3), 2396–2406. https://doi.org/10.1523/jneurosci.15-03-02396.1995.

26. Needleman, Saul B., and Christian D. Wunsch. (1970) A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins, *Journal of Molecular Biology.* 48 (3), 443–453. https://doi.org/10.1016/0022-2836(70)90057-4.

27. Lee, J., Patel, D. S., Ståhle, J., Park, S.-J., Kern, N. R., Kim, S., Lee, J., Cheng, X., Valvano, M. A., Holst, O., Knirel, Y. A., Qi, Y., Jo, S., Klauda, J. B., Widmalm, G., and Im, W. (2018) CHARMM-GUI Membrane Builderfor Complex Biological Membrane Simulations with Glycolipids and Lipoglycans, *J. Chem. Theory Comput.* 15 (1), 775–786. *https://doi.org/10.1021/acs.jctc.8b01066*

28. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys.* 79 (2), 926–935. https://doi.org/10.1063/1.445869

29. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys.* 79 (2), 926–935. https://doi.org/10.1063/1.445869

30. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable Molecular Dynamics with NAMD, *J. Comput. Chem.* 26 (16), 1781–1802. *https://doi.org/10.1002/jcc.20289*

31. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., and MacKerell, A. D. (2016) CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins, *Nat. Methods* 14 (1), 71–73. *https://doi.org/10.1038/nmeth.4067*

32. Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A. D., and Pastor, R. W. (2010) Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types, *J. Phys. Chem. B* 114 (23), 7830–7843. *https://doi.org/10.1021/jp101759q*

33. Miyamoto, S., and Kollman, P. A. (1992) Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models, *J. Comput. Chem.*13 (8), 952–962. https://doi.org/10.1002/jcc.540130805

34. Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K., and Granier, S. (2012) Crystal Structure of the Mu-Opioid Receptor Bound to a Morphinan Antagonist. *https://doi.org/10.2210/pdb4dkl/pdb*

35. Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J. Chem.Theory Computat.* 9 (7), 3084–3095. *https://doi.org/10.1021/ct400341p*

36. Ricarte, A., Dalton, J. A. R., and Giraldo, J. (2021) Structural Assessment of Agonist Efficacy in the μ-Opioid Receptor: Morphine and Fentanyl Elicit Different Activation Patterns, *J. Chem. Inf. Model.* 61 (3), 1251–1274. *https://doi.org/10.1021/acs.jcim.0c00890*

37. Cong, X., Campomanes, P., Kless, A., Schapitz, I., Wagener, M., Koch, T., and Carloni, P. (2015) Structural Determinants for the Binding of Morphinan Agonists to the μ-Opioid Receptor, *PLOS ONE* 10 (8), e0135998. *https://doi.org/10.1371/journal.pone.0135998*

38. Lipiński, P. F. J., Jarończyk, M., Dobrowolski, J. Cz., and Sadlej, J. (2019) Molecular Dynamics of Fentanyl Bound to μ-Opioid Receptor, *J. Mol. Model.* 25 (5). https://doi.org/10.1007/s00894-019-3999-2.

39. Gimpl, G. (2016) Interaction of G Protein Coupled Receptors and Cholesterol, *Chem. Phys.Lipids* 199, 61–73. *https://doi.org/10.1016/j.chemphyslip.2016.04.006*

40. Vukojevic, V., Ming, Y., D'Addario, C., Hansen, M., Langel, Ü., Schulz, R., Johansson, B., Rigler, R., and Terenius, L. (2008) μ-Opioid Receptor Activation in Live Cells, *FASEB J.* 22 (10), 3537–3548. *https://doi.org/10.1096/fj.08-108894*

41. Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021) AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings, *J. Chem. Inf. Model.* 61 (8), 3891–3898 *https://doi.org/10.1021/acs.jcim.1c00203*

42. Trott, O., and Olson, A. J. (2009) AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 31 (2), NA-NA. *https://doi.org/10.1002/jcc.21334*

43. Vanommeslaeghe, and K., MacKerell, A. D. (2012) Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing, *J. Chem. Inf. Model.* 52 (12), 3144–3154. *https://doi.org/10.1021/ci300363c*

44. Vanommeslaeghe, K., Raman, E. P., and MacKerell, A. D. (2012) Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges, *J. Chem. Inf. Model.* 52 (12), 3155–3168. *https://doi.org/10.1021/ci3003649*

45. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell, A. D. (2010) CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields, *J. Comput. Chem.* 31 (4), 671–690. *https://doi.org/10.1002/jcc.21367*

46. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics, *J. Mol. Graph.* 14 (1), 33–38. *https://doi.org/10.1016/0263-7855(96)00018-5*

47. Mayne, C. G., Saam, J., Schulten, K., Tajkhorshid, E., and Gumbart, J. C. (2013) Rapid Parameterization of Small Molecules Using the Force Field Toolkit, *J. Comput. Chem.* 34 (32), 2757–2770. *https://doi.org/10.1002/jcc.23422*

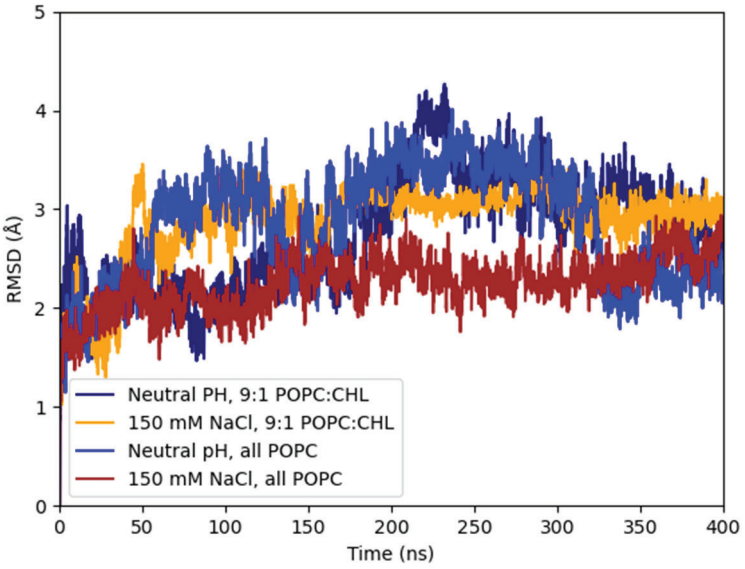
ABOUT THE STUDENT AUTHORS

Allison Barkdull is a senior undergraduate at the University of Florida studying biomedical engineering with a minor in physics. Lexin Chen is a chemistry Ph.D. student at the University of Florida. Her research is on building algorithms to analyze the molecular similarity of large datasets with applications to drug discovery. Akash Mathavan is a fourth year M.D. candidate at the University of Florida College of Medicine pursuing a career in internal medicine.

PRESS SUMMARY

The mu-opioid receptor is the primary target for pain-modulating drugs. However, existing opiate drugs have detrimental side effects such as addiction and respiratory depression. Molecular dynamics simulations allow the motion of the mu-opioid receptor to be studied with atomistic resolution. However, because the mu-opioid receptor is enclosed within the lipid bilayer, molecular dynamics simulations typically require the simulation of lipid molecules as well as the receptor, which requires a large amount of computational resources. This study proposes and evaluates an alternative method to simulate the mu-opioid receptor by adding restraints on the transmembrane regions of the protein to mimic the viscosity of the lipid bilayer without explicitly simulating lipid molecules. The proposed model allows the dynamics of the mu-opioid receptor to be simulated 49% faster than a model that involves explicitly simulated lipid molecules. This has the potential to speed up studies of mu-opioid receptor motion, aiding in discovering new opioid drugs.

SUPPLEMENTAL INFORMATION



Supplemental Figure 1. The RMSD of MOR in an explicitly modeled lipid bilayer with either a membrane of either 100% POPC or 10% cholesterol and 90% POPC.

A Review of Models on Direct Evaporative Cooling

*Michael Wilkins & Nelson Fumo**

Department of Mechanical Engineering, University of Texas at Tyler, Tyler, TX

https://doi.org/10.33697/ajur.2023.086

Students: mwilkins5@patriots.uttyler.edu, mtwilkins903@gmail.com
*Mentor: nfumo@uttyler.edu**

ABSTRACT
Direct evaporative cooling (DEC) is a technology that is continuously expanding into different areas of study. The foundation of this process has been built through expansive research efforts and physical experimental data. The ability to accurately model and predict the performance of DEC systems allows the energy-efficient process to gain traction in HVAC applications, however, the inconsistencies present among research efforts created discontinuities in the reproduction of a system. By reviewing current literature, the discrepancies in the defining methodologies of how DEC systems are defined and predicted can provide insight to future research. This review depicts the different approaches taken in recent research to define the equations that govern the thermodynamic processes, the different materials used in the process, and the models used to predict the performance of DEC systems. By identifying the most common practices in current research, the gaps in literature can be recognized and overcome in further efforts.

KEYWORDS
Direct Evaporative Cooling; Evaporative Cooler; Evaporative Cooling Media; HVAC; Cooling Effectiveness

INTRODUCTION

The ever-growing issue of increasing global energy consumption and CO_2 emissions[1] demands a further review of pertinent solutions. Throughout the years, since its invention, air conditioning systems have become an integral part of most commercial buildings and households, which are responsible for a considerable amount of the total energy consumed worldwide. The U.S Department of Energy states that roughly 6% of all supplied energy is consumed by household air conditioners, with even larger numbers corresponding to commercial building HVAC systems[2]. The necessity to improve efficiency and decrease power consumption related to HVAC systems is crucial to the sustainability of the future. Evaporative cooling (EC) has the potential to help HVAC systems achieve these objectives.

EC is a fairly simple process that is capable of producing economical cooling when weather conditions are appropriate, an aspect of which is part of the limitations preventing the larger market penetration of evaporative cooling technology. While evaporative coolers are known to achieve significant temperature drops in more arid climates, they are also known they can provide relief from the heat in any climate with high temperatures. The evaluation of similar EC processes has yet to gain uniformity among available research, which causes discontinuities in the progression of the technology over time. To further research efforts and optimize this process for expanded use in the HVAC industry, there must be common methods of practice for defining EC systems and their associated performance parameters. In doing so, the uncertainty involved in research can be addressed and overcome, leading to a faster progression of the advancement of evaporative cooling technologies. The appropriate design and optimization of EC systems relies on the ability of accurate models to reproduce how the system will perform. This study focuses on models for the evaporative cooling process that occurs at the media of an EC system. As shown in **Figure 1**, the first component of a basic DEC system is a media that allows for an increase in the contact, and therefore heat and mass transfer, between the air to be conditioned and the water used to condition it. Another component includes a water recirculation subsystem in which a pump takes water from a water sump and supplies it to the water distributor that will uniformly distribute the water over the media. The last component is a fan that moves the air through the media.

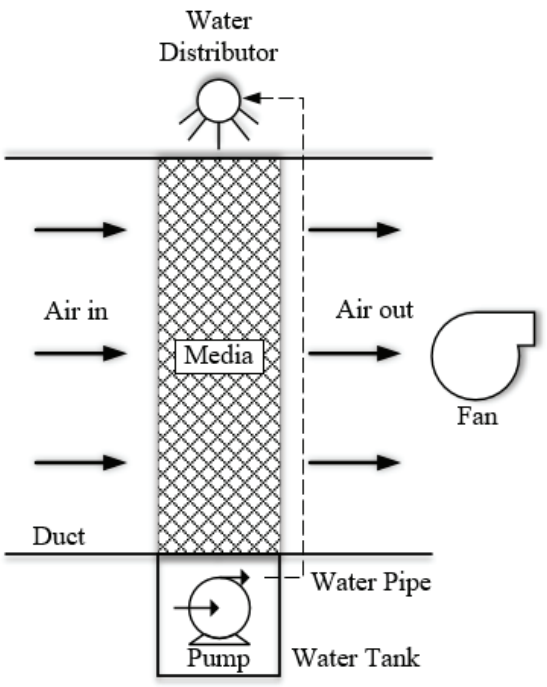


Figure 1. Basic layout of a DEC cooler.

Every component of the EC system has an impact on the overall performance of the EC system, whose objective is to lower the temperature of incoming air by increasing its moisture content in an adiabatic process in which the air is cooled at constant wet-bulb temperature. This is achieved by transferring heat from the air to the water, which causes an evaporation effect as the air flows through the water-saturated media[3]. The water contained in the sump is delivered to the system with the intention of fully saturating the media; however, full saturation is influenced by each component. The media's ability to saturate completely is referred to as "saturation efficiency". This concept and the associated issues are discussed in further detail in *Water Distribution Methods*. Since the incoming air stream will be cooled at a constant wet-bulb temperature, the maximum attainable temperature drop is determined by the initial dry bulb temperature and the humidity of the incoming air as seen in **Figure 2**. The process moves the state of the air from "inlet" to "outlet" along the wet-bulb line, which creates the assumption that no other methods of heat transfer are affecting the airflow.

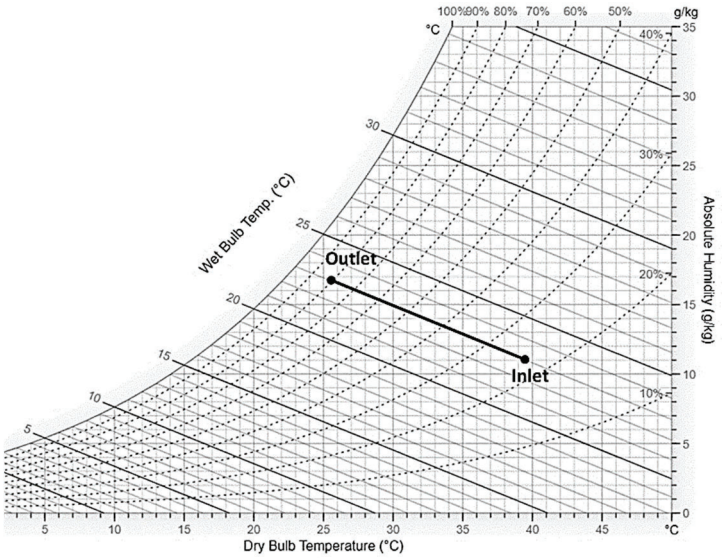


Figure 2. Psychometric process of DEC.

The methods of evaluating and predicting the performance of EC systems often exclude the layout of the EC system being analyzed. The inclusion of stating the physical layout of the analyzed systems is imperative to fully define the characteristics of each system. Furthermore, the equations to determine the efficiency of these systems vary among previous research efforts. This creates inconsistencies in the continual improvement of the involvement of EC systems in various applications.

The performance evaluation of EC systems is determined by design parameters affecting the thermodynamic process. How the evaluation process is described depends on the type of approach or model used, which is reviewed and presented in this paper. There are many ways to characterize EC systems based on the components and conditions that make up the systems, however there are limited resources that help to distinguish how these characteristics perform given the application. The ability to model the operating conditions of each component of an EC system would provide insight into the design of the components and layout of these systems. This review serves the purpose of providing a comparison of systems and applications with focus on the models for media performance, as a means of defining characteristics of current research efforts on DEC systems.

THERMODYNAMIC ANALYSIS OF DEC SYSTEMS

The performance of a DEC system is determined by the ability to modify the thermodynamic properties of the air as it exits the system. The measure of efficiency may be approached in different ways, therefore uniformity among presented results should be considered based on the method of analysis of the system. In **Table 1**, the various equations used to determine the efficiency are represented, along with the researchers that support the equation.

	Property	Equation	Reference
COOLING EFFICIENCY	Temperature	$\varepsilon = \frac{T_{inlet} - T_{outlet}}{T_{inlet} - T_{wb}}$ (1)	**4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17**
	Humidity Ratio	$\varepsilon = \frac{\Delta\omega}{\Delta\omega_{maximum}} = \frac{\omega_{inlet} - \omega_{outlet}}{\omega_{inlet} - \omega_{maximum}}$ (2)	**18, 19**
	Enthalpy	$\varepsilon = \frac{h_{outlet} - h_{inlet}}{h_{wb,outlet} - h_{inlet}}$ (3)	**20**
	Exergy	$\varepsilon = \frac{h_{inlet} - h_{outlet} - T_{inlet}(s_{inlet} - s_{outlet})}{h_{inlet} - h_{wb,inlet} - T_{inlet}(s_{inlet} - s_{wb,i})}$ (4)	**1**

Table 1. Efficiency equations and references.

As seen from **Table 1**, the most common method of determining the efficiency of a DEC is based on temperature. For this case, the performance is better called effectiveness as it compares the inlet and outlet air dry-bulb temperatures with the maximum possible change in air dry-bulb temperature, which is also called wet-bulb depression. This approach only considers the effects on the airflow and does not state the quality of the supplied water to the system. As a result of this, for a specific airflow rate and air inlet condition, the DEC system may perform inconsistently based on the inlet temperature of the water being supplied to the system. The humidity ratio approach is less common throughout research. This may be due to methods of measuring humidity, as well as its associated accuracy, in comparison to the simplicity of temperature measurements. The psychometric performance evaluation of the system using the enthalpy approach considers temperature and humidity of the airflow for a more extensive approach in determining the efficiency of the DEC system. The exergy approach, stated in the review by S. R. Pinar Mert Cuce et al[1], follows the same ideology as the enthalpy approach, but adds an exergy destruction term. Exergy destruction is directly related to the second law of thermodynamics and describes the system's irreversible losses. This provides an added element of reliability to the statement of the system's operation and efficiency.

Coefficient of Performance

The coefficient of performance of a DEC thermodynamic process relates to the amount of cooling from the evaporative cooling effect and the power used by the system, which is associated to the fan and water pump. For EC systems, the input energy to system typically is associated to the pump work and the fan work; however, the standard form of the equation applies to all the research included in this review, as seen in **Equation 1**.

$$COP = \frac{Q_{c,e}}{\Sigma W} \quad \textbf{Equation 1.}$$

For the COP equation, $Q_{c,e}$ represents the effective heat exchange performed by the system and ΣW is the summation of all the work inputs to the system. As can be noticed, the *COP* is a measure of the system performance and the effectiveness is a measure

of the efficiency of the psychrometric process, which defines the cooling capacity. As expected, there was no difference in how the *COP* was determined among authors.

Input Parameters

To determine a DEC system's performance, the measured and controlled variables that describe the inlet conditions of the system must be defined. The various subsystems of a DEC system can be divided into three main areas: airflow, water distribution, and EC media. To accurately define the state of the airflow entering and exiting a DEC system, the psychometric properties and the flow rate must be stated. As seen in **Table 2**, the defining factors used in research can be separated by the airflow's input variables.

	Parameters	Symbols	Reference
INPUT PARAMETERS	Temperature, Humidity Ratio and Velocity	$T_{inlet}, \omega, v_{inlet}$	**4, 5, 13, 14, 17, 18, 22**
	Temperature, Humidity Ratio, and Volumetric Flow Rate	$T_{inlet}, \omega, \dot{V}_{inlet}$	**8, 15**
	Temperature, Velocity, and Molar Concentration	$T_{inlet}, v_{inlet}, c_{a,in}$	**9, 11**
	Temperature, Humidity Ratio, and Mass Flow Rate	$T_{inlet}, \omega, \dot{m}_a$	**10, 12**
	Temperature, Humidity Ratio, Pressure, and Specific Enthalpy	$T_{inlet}, \omega, P, h_a$	**20**

Table 2. Input parameters of DEC systems.

Water Distribution

The water distribution subsystem consists of many different components that impact the performance of a DEC system. This includes the supply water, the pump, and the method of distributing the water. The defining qualities of the water as part of a DEC system are the temperature and the mass flow rate. Although well know, the associated formula for pump work is given to illustrate its consistent use among researchers and is shown in **Equation 2**.

$$W = \frac{Q_w \rho_w H_p g}{\eta_{pump}} \qquad \textbf{Equation 2.}$$

In **Equation 2**, Q_w represents the volumetric flow rate of the water, ρ_w describes the water's density, H_p is the total head, η_{pump} is the total efficiency of the pump, and g is gravity.

EC Media Input Characteristics

The performance of the media used in DEC systems is also described by many different parameters which, as stated before, creates a divide in how the DEC media is classified among researchers. **Table 3** shows the various terminology and data used to describe a material's effectiveness.

	Terminology	Reference
EFFECTIVENESS	Saturation Efficiency	**22**
	Unity Wettability Factor	**6**
	Cooling Capacity	**8**
	Heat Transfer Coefficient	**9**
	Mass Transfer Coefficient	**9**
	Water Absorption Capacity	**10**

Table 3. Effectiveness of DEC media.

It should be stated that the differences in terminology of the effectiveness of EC media primarily stems from the goal of the research. However, the various methods of describing the performance lack consistency and uniformity in the research reviewed.

EVAPORATIVE COOLING MEDIA

The main component of the DEC system is the media that defines the area of heat and mass transfer for the air flowing through the media, resulting in a reduction in air temperature in conjunction with an increase in humidity. The evaporative media can be broken down into certain controlled characteristics that impact a DEC system's efficiency:

1. Material
2. Cross Sectional Area
3. Thickness
4. Shape

Each of these characteristics play a role in how the media performs based on the given application. Manipulation of these characteristics can result in an increase in efficiency; however, the system may be penalized by an increase in undesirable qualities such as pressure drop. The optimal arrangement of these variables can allow the system to achieve its purpose with minimal negative qualities.

The DEC pad's material and shape affect performance parameters such as saturation capacity and pressure drop. Losses in efficiency are introduced in many ways, such as when the material is unable to saturate completely and requires a greater flow rate to keep the contact surfaces moist. The shape and size of the material also introduces losses to the system by increasing the pressure drop. Pressure drop decreases the overall efficiency of the system by requiring more power to achieve the desired rate of airflow through the system. Of the four characteristics of DEC pads, the material has a great influence on the others due to material constraints and manufacturing abilities. The materials found to be used in the literature reviewed for this study are further analyzed and compared in the following section.

Material:
Evaporative cooling pads can be broken down into the three main categories: rigid pads, fiber pads, and package/fill pads[3]. The primary reason for experimentation with different materials or types of pads is to attempt to find materials with greater saturation capacity, which correlates to heat and mass transfer to the airflow across the media.

Rigid Pads:
Rigid pads most commonly consist of a series of individual sections of material, which are formed to a specific shape to optimize surface area and assembled with an adhesive of joining process that affect the materials pressure drop and air flow characteristics. **Figure 3** shows the relative geometric properties of this type of pad.

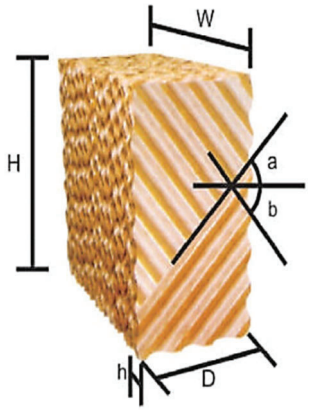


Figure 3. Geometric properties of rigid pads[21]

A major benefit of rigid pads is the inherent ability to design shape profiles for a specific application. As the most used rigid media currently being researched, CELdek has numerous studies on sizing systems based on airflow requirements. **Figure 4** represents the performance of this type of rigid pad comparing different thicknesses, over a wide range of flowrates.

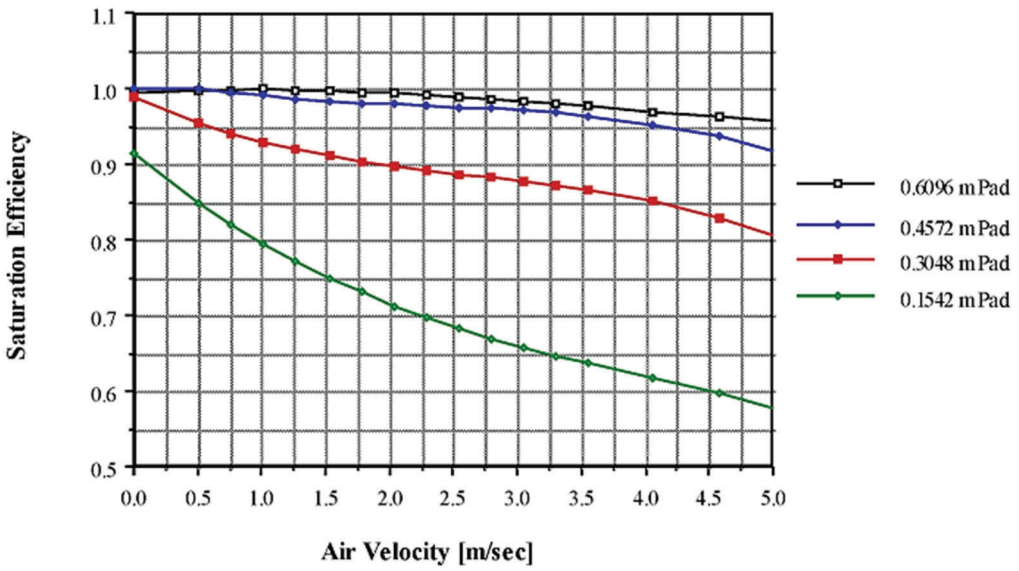


Figure 4. Saturation efficiency vs. pad thickness of Celdek DEC pads[23]

Figure 4 follows the curve fit equation represented in **Equation 3**[23], with input variables being the thickness, depth of the pad, and the incoming air velocity.

$$\varepsilon = 0.792714 + 0.958569t - 0.25193\text{v} - 1.03215t^2 + 0.0262659v^2 + 0.914869(t * v) - 1.48241(v * t^2) - 0.018992(v^3 * \text{t}) + 1.13137(t^3 * v) + 0.0327622\,(v^3 * t^2\,) - 0.145384(v^2 * t^3\,)$$

Equation 3.

Fiber Pads

Fiber pads primarily consist of organic or synthetic materials that are woven together and joined using an adhesive or a joining process. These pads have attributes that can be controlled through the manufacturing process, such as thickness, cross-sectional area and spacing of the fibers. This allows the designer to control characteristics such as the pad's potential for saturation capacity, pressure drop and airflow capacity. Common materials used in fiber pads include vegetable fibers, textile fibers or woven fabrics, paper, wood, plastic, or stone[3]. A common fiber pad used in the industry is produced by AirCare and is shown in **Figure 5**.

Figure 5. Fiber pad produced by AirCare[31].

It should be noted that only one of the researchers included in this review generated a predictive model using this material as a DEC media[17].

Package/Fill Pads

Package pads, otherwise known as fill pads, use less structured materials contained within a case to distribute the flow through the media. The media used to fill the case usually consists of porous and inorganic materials, such as volcanic stones or expanded clay. The casing material is constructed with plastic or metal mesh. This limits the restrictive qualities contained with most media pads and allows flow through the structure depending on the type of media contained within the casing[3]. This type of pad lacks manufactured options within the market and no researchers in this review provided predictive modeling using this specific type of DEC media pad.

DIRECT EVAPORATIVE COOLING PAD MODELS

The implications of DEC systems are expanding as research continues and advancements in manufacturing capabilities increase. The type of EC model used in each application has the potential to alter the determination on if the DEC system is a viable enhancement or a detriment. The use of desiccant systems may be beneficial for systems utilizing rigid media pads, however, may not have a significant effect on other types of media pads. Through a review of the many different applications of DEC systems in current research, the progress of DEC in the future of HVAC and efficient heat exchange systems may be optimized.

As a starting point to discuss the models for analyzing DEC, the authors want to clarify that, when possible, the models were classified as data-driven, engineering, or hybrid models. The data-driven models are those that use any type of statistical approach such as regression analysis, artificial intelligence such as neural networks, or support vector machines among others. The engineering models use physical principles describing the natural phenomena affecting system behavior. The hybrid or grey models are a combination of the previous two, which are developed when the physical information of the system does not allow defining the system completely and the parameters needed for a full description of the system are obtained by statistical analysis. As for any model that is not developed from a pure analytical analysis, experimental data is needed for validation. In this sense, it is important to consider the **ANSI/ASHRAE Standard 133-2015** Method of Testing Direct Evaporative Air Coolers[24] as a reference when performing experiments. This is particularly important regarding accuracy for air parameters such as temperatures, air flow, water flow, and pressure drop.

Data Driven Models:

Data driven models are those formulated based strictly on experimental data obtained from physical testing. These models are often dependent on the characteristics of the physical setup associated with the experiment, therefore should be carefully considered when conducting further research.

The Engineering Reference[25] of the whole building energy simulation program **EnergyPlus** [26] uses a multi-variate least squares curve fit to estimate the saturation efficiency (ε)as a function pad face velocity (*Airvel*) and pad thickness (*Depth*). The coefficients of the third order quadratic curve fit shown in **Equation 4** [25] were obtained using data from the manufacturer of the CelDek rigid media pad.

$$\begin{aligned}\varepsilon = {} & 0.792714 + 0.958569(Depth) - 0.25193(Airvel) - 1.03215(Depth^2) + 0.0262659(Airvel^2) \\ & + 0.914869(Depth \times Airvel) - 1.48241(Airvel \times Depth^2) \\ & - 0.018992(Airvel^3 \times Depth) + 1.13137(Depth^3 \times Airvel) \\ & + 0.0327622(Airvel^3 \times Depth^2) - 0.145384(Depth^3 \times Airvel^2)\end{aligned} \quad \textbf{Equation 4.}$$

In the experimental process performed by M.C. Ndukwu et al.[17], a modeling equation was used which was based on physical data for the different materials used in the experiment. The variables included in the model were inlet air temperature, wet bulb temperature, relative humidity, mass flow rate of the air, and volume of the humidifier as shown in **Equations 5-7**. The output of the equation is the outlet air temperature of the DEC system.

Jute Fiber: $R^2 = 0.81$

$$T_c = T_{ab} - [(T_{ab} - T_w)(0.226 - 0.0094\omega + 2.55\dot{m}_a + 0.00203V_H)] \quad \textbf{Equation 5.}$$

Palm Fruit Mesocarp Fiber: $R^2 = 0.78$

$$T_c = T_{ab} - [(T_{ab} - T_w)(0.182 + 0.0131\omega - 0.0546\dot{m}_a + 0.0016355V_H)] \quad \textbf{Equation 6.}$$

Wood Charcoal: $R^2 = 0.76$

$$T_c = T_{ab} - [(T_{ab} - T_w)(1.52 - 0.006288\omega - 1.739\dot{m}_a - 0.00912V_H)] \quad \textbf{Equation 7.}$$

In a physical experiment performed by S.A. Nada et al.[20], heat and mass transfer characteristics of corrugated cellulose papers were evaluated and modeled using data obtained from an experimental setup at Benha University (Egypt). Regressions were formulated and documented with the associated error ranges associated. The formulas use the inlet air temperature, water temperature, mass flow rate of the water, the evaporative pad thickness(δ), Reynold's number, and Prandtl's number. The equations for the outlet air temperature and the relative humidity are shown in **Equations 8-9**.

$$t_{a,out}(℃) = 21.2\left(\frac{t_{a,in}}{t_{a,Ref}}\right)^{0.752}\left(\frac{t_w}{t_{w,Ref}}\right)^{1.275}\left(\frac{\dot{m}_w}{\dot{m}_{w,Ref}}\right)^{-0.0313}\left(\frac{\delta}{\delta_{Ref}}\right)^{-0.099} Re^{0.133}Pr^{0.33} \quad \textbf{Equation 8.}$$

$$RH_{out}(\%) = 200.1\left(\frac{t_{a,in}}{t_{a,Ref}}\right)^{0.298}\left(\frac{t_w}{t_{w,Ref}}\right)^{0.405}\left(\frac{\dot{m}_w}{\dot{m}_{w,Ref}}\right)^{0.145}\left(\frac{\delta}{\delta_{Ref}}\right)^{0.16} Re^{-0.088}Pr^{0.33} \quad \textbf{Equation 9.}$$

Engineering Models

Throughout the research, the vast majority of engineering models were all validated by previously performed physical experimentation. There was a lack of predictive modeling based on mathematical concepts alone.

Hybrid Models

In the research regarding parameters affecting direct and indirect evaporative cooling systems[27], the authors state that the data corresponds to experiments conducted at BZU Multan (Pakistan) and references for the detailed information are given.

The surrogate model proposed by Hussain et al.[27] uses meteorological data (dry bulb temperature, dew point temperature, wet bulb temperature, relative humidity, enthalpy, and humidity ratio) and system parameters (area and inlet velocity) as inputs to a Gaussian process that feeds a neural network to predict system performance.

Research performed by Qi Zhang et al.[11] describes a numerical model that is a tested and validated expansion of the physical experimentation tested by Yan et al.[28]. The authors use computational geometry based on the EC media "CELdek 7060", which is implemented with associated equations into a finite element software. The simulation follows logarithmic distribution laws for the temperature layers of grid node layouts. A SIMPLE algorithm is used for the analysis and the turbulent model is consistent with a turbulent kinetic energy(ϰ) and dissipation rate(ε) for the control volume and adjacent areas. It was reported that a 0.32℃ error differential was calculated between the physical and numerical models.

A performance analysis was performed by Xin Cui et al.[9] as a numerical validation of previous research performed by Xiangjie Chen et al.[29]. The membrane-based DEC system was modeled in the COMSOL Multiphysics environment and solved by using three physical models: the heat transfer mode, the laminar flow mode, and the transport of diluted species mode. The temperature, concentration, and the velocity fields of the domain were solved simultaneously in the simulation.

Research on the application of hollow fiber membranes in EC systems was performed and modeled by Weichao Yan et al.[32]. As a nontraditional form of evaporative cooling, the typical dimensions of the evaporative cooling media are replaced with descriptive dimensions of the membrane being used in the experimental process. This creates additional input parameters to the predictive modeling shown in outlet temperature and relative humidity **Equations 10 - 11**. The equations also have associated parameters outlining the constraints of each element within the applicable range of modeling within the research paper. The parameters in **Equations 10 - 11** are temperature (*T*), relative humidity (*RH*), velocity (υ), length (*L*), diameter (*d*), thickness (δ), packing fraction (φ), and water – air ratio (μ).

$$T_{out} = -0.6204 + 0.612T_{in} + 7.049\omega_{in} + 1.922v - 1.660\mu - 13.094L + 10.830d_i + 12.888\delta - 51.065\varphi + (0.476T_{in}\omega_{in}) + (0.024T_{in}v) - (0.787T_{in}\varphi) - (2.012\omega_{in}v) + (13.787\omega_{in}L) - (6.903\omega_{in}d_i) + (71.092\omega_{in}\psi) - (0.769v\mu) - (3.926vL) + (2.135vd_i) + (8.263v\delta) - (16.638v\varphi) - (12.490Ld_i) - (50.080L\delta) + (83.675L\varphi) - (55.262d_i\varphi) - (5.246{\omega_{in}}^2) + (3.255\mu^2) + (22.738L^2) + (348.796\varphi^2)(℃)$$ **Equation 10.**

$$RH_{out} = 0.415 + 0.008T_{in} + 1.518\omega_{in} - 0.2329v - 0.027\mu + 0.316L - 0.404d_i - 5.147\delta + 4.439\varphi + (0.062T_{in} * \delta) - (0.201\omega_{in} * v) - (1.155\omega_{in} * L) + (0.188\omega_{in} * d_i) + (3.983\omega_{in}\delta) - (4.285\omega_{in}\varphi) + (0.123vL) - (0.077vd_i) + (0.601v\varphi) + (0.372Ld_i) - (2.038L\varphi) + (2.004d_i\varphi) - \left(0.0001{T_{in}}^2\right) - (1.272{\omega_{in}}^2) + (0.013v^2) - (1.276L^2) - (16.070\varphi^2)$$ **Equation 11.**

Research performed by M. Lata et al.[30] outlines a finite difference method utilizing MATLAB. The heat and mass transfer approach implements energy balance equations for the system. The system variables used in the simulation are mass flow of air, mass flow of water, thickness and surface of cooling pad, and the inlet temperature conditions of both the air and water. The mathematical model was validated using experimental data obtained by the authors in 2018 in Ahmedabad, Gujarat, India.

Physical Model Experimental Structure

The literature reviewed in this research is displayed in **Table 4** by analyzing the application, in conjunction with the type of DEC pad material, water delivery system, and the size of the media pad.

The idea of this section is to refer and briefly describe the different types/arrangements found as source of models. For example, from the first set of papers:

MODEL	EVAPORATIVE MEDIA	WATER DELIVERY	MEDIA SIZE
Ground source direct evaporative cooling[5]	Not Stated	Pulverized nozzles	Not Stated
Evaporative cooling with pre-dehumidification with activated carbon fiber[4]	GLASdek7090	Spray nozzles	500 mm × 400 mm × 300 mm
Evaporative cooling with post-dehumidification Aluminophosphate zeolite[18]	Cellulose coated with AlPo zeolite desiccant	Spray nozzle	200 mm x 200 mm x 198 mm
Regenerative Evaporative cooling[6]	Kraft Paper	Not stated	1200mm x 80mm x 5 mm gap thickness
Various Cross-Sectional Metal Sheet Shapes[22]	Aluminum	Perforated distribution tube/pipe	700mm x 700mm x 200mm
Desiccant assisted cooling systems[10]	Wood Chips, Yellow Stone, Pumice, Eucalyptus Fibers, Vermiculite	Spray Nozzle	240mm Diameter x 50mm
Peformance Enhancement of DE builidng[20]	Corrugated Cellulose Fibers, "beehive" structure	Steel pipe with evenly spaced holes	335mm x 390mm x (35,70,105,140) mm
Numerical Simulation[11]	CELdek7060	Distribution Pipe	300mm x 200mm x 28mm

Table 4. Models of DEC systems.

DISCUSSION & CONCLUSION

The purpose of this research was to review the performance models of DEC systems. The review of this material allows for the methodology of each researcher to be compared based on similarities and differences. The defining characteristics of the DEC systems, the setups and types of media, and methodology of model generation all impact the cohesiveness of research efforts. As discussed, the equations defining cooling efficiency as well as the input parameters used in various literature lack consistency, which limits the amount of comparable data for the different processes. The approach taken by researchers, as stated in the literature, was to describe processes in a manner that was deemed suitable per given application, which resulted in widespread variance. The "hybrid" direct evaporative cooling pad models outnumbered the amount of solely data driven models and engineering models. This states that the majority of researchers use experimental data to either validate the performance models or aid in the derivation of a mathematical model. Throughout the reviewed literature, there are numerous researchers testing various DEC topologies and media, however none of them provided modeling for performance based on the described parameters. This study shows the need of the search of general accepted method for development of models for DEC systems, which prevalence is toward hybrid models.

ACKNOWLEDGEMENTS
The authors express gratitude for the support received from the ASHRAE East Texas Chapter in the form of a scholarship.

Nomenclature

COP	*Coefficient of Performance*	ε	*Effectiveness*
DEC	*Direct Evaporative Cooling*	ϱ	*Density*
EC	*Evaporative Cooler*	*h*	*Enthalpy*
T	*Temperature (°C)*	$\dot{W}_p$	*Pump Work*
P	*Pressure*	$\dot{W}_f$	*Fan Work*
RH	*Relative Humidity*	*wb*	*Wet bulb*
C_{pa}	*Specific Heat of Air*		
$\dot{m}$	*Mass Flow Rate*		

REFERENCES

1. S. R. Pinar Mert Cuce. (2016) A state of the art review of evaporative cooling systems for building applications, *Renewable and Sustainable Energy Reviews* 54, 1240-1249. *https://doi.org/10.1016/j.rser.2015.10.066*
2. U.S. Department of Energy, *Air Conditioning*, *https://www.energy.gov/energysaver/air-conditioning* (accessed Nov 2022)
3. A. Tejero-González, A. Franco-Salas. (2021) Optimal operation of evaporative cooling pads: A review, *Renewable and Sustainable Energy Reviews* 151, no. 111632. *https://doi.org/10.1016/j.rser.2021.111632*
4. Bianfeng Yang, Cong Wang, Xu Ji, Ren Zhang, Xirong Yue, Junneng Nie. (2022) Two-stage evaporative cooling system with composite activated carbon fiber dehumidification, *International Journal of Thermal Sciences* 179, no. 107725.
5. Gokhan Yıldız, Alper Ergün, Ali Etem Gürel, Ümit Agbulut, Servet Eser, Asif Afzal, C.Ahamed Saleel. (2022) Exergy, sustainability and performance analysis of ground source direct evaporative cooling system, *Case Studies in Thermal Engineering* 31, no. 101810. *https://doi.org/10.1016/j.csite.2022.101810*
6. Sarvesh Kashyap, Jahar Sarkar, Amitesh Kumar. (2022) Comparative performance analysis of different novel regenerative evaporative cooling device topologies, *Applied Thermal Engineering* 176, no. 115474.
7. Manju Lata, Dileep Kumar Gupta. (2020) Experimental Investigation and Simulation of Modified Evaporative Cooling System, *Advances in Mechanical Engineering*, 441-448. *https://doi.org/10.1007/978-981-15-3639-7_52*
8. Weichao Yan, Xiangzhao Meng, Xin Cui, Yilin Liu, Qian Chen, Liwen Jin. (2022) Evaporative cooling performance prediction and multi-objective optimization for hollow fiber membrane module using response surface methodology, *Applied Energy* 325, no. 119855. *https://doi.org/10.1016/j.apenergy.2022.119855*
9. Xiangjie Chen, Yuehong Su, Devrim Aydin, Xingxing Zhang, Yate Ding, David Reay, Richard Law, Saffa Riffat. (2020) Performance analysis of a hollow fiber membrane-based heat and mass exchanger for evaporative cooling, *Applied Energy* 271, no. 115238. *https://doi.org/10.1016/j.apenergy.2020.115238*
10. Nima Khosravi, Devrim Aydin, Majid Karim Nejhad, Pervin Abohorlu Dogramaci. (2020) Comparative performance analysis of direct and desiccant assisted evaporative cooling systems using novel candidate materials, *Energy Conversion and Management* 221, no. 113167. *https://doi.org/10.1016/j.enconman.2020.113167*
11. Qi Zhang, Suoying He, Jun Cheng, Bin Zhao, Xuehong Wu, Mingxuan Yan, Ming Gao, Zhe Geng, Shuzhen Zhang. (2022) Numerical simulation of evaporative cooling process in a medium-gap-medium arrangement, *International Journal of Thermal Sciences* 179, no. 107700. *https://doi.org/10.1016/j.ijthermalsci.2022.107700*
12. Rekabra Youssef, Md Sazzad Hosen, Jiacheng He, Joris Jaguemont, Lysander De Sutter, Joeri Van Mierlo, Maitane Berecibar. (2021) Effect analysis on performance enhancement of a novel and environmental evaporative cooling system for lithium-ion battery applications, *Journal of Energy Storage* 37, no. 102475. *https://doi.org/10.1016/j.est.2021.102475*
13. Ali Hammoodi Mahdi, Issam M. Ali Aljubury (2021) Experimental investigation of two-stage evaporative cooler powered by photovoltaic panels using underground water, *Journal of Building Engineering* 44. *https://doi.org/10.1016/j.jobe.2021.102679*
14. Shiva Kumar, Sampath Suranjan Salins, S.V. Kota Reddy, Prasanth Sreekumar Nair (2021) Comparative performance analysis of a static & dynamic evaporative cooling pads for varied climatic conditions, *Energy* 223, no. 101136. *https://doi.org/10.1016/j.energy.2021.121136*
15. Ibrahim Atmaca, Ali Senol, Ahmet Çaglar. (2022) Performance testing and optimization of a split-type air conditioner with evaporatively-cooled condenser, *Engineering Science and Technology* 32, no. 101064. *https://doi.org/10.1016/j.jestch.2021.09.010*
16. Sujatha Abaranji, Karthik Panchabikesan, Velraj Ramalingam. (2021) Experimental study on the direct evaporative air-cooling system with vermicompost material as the water storage medium, *Sustainable Cities and Society* 71, no. 102991. *https://doi.org/10.1016/j.scs.2021.102991*
17. M.C.Ndukwu, M.I.Ibeh, E.C.Ugwu, D.O.Igbojionu, A.A.Ahiakwo, Hongwei Wu. (2022) Evaluating coefficient of performance and rate of moisture loss of some biomass humidifiers materials with a developed simple direct stand-alone evaporative cooling system for farmers, *Energy Nexus* 8, no. 100146. *https://doi.org/10.1016/j.nexus.2022.100146*

18. Seng Sunhor, Dendi Adi Saputra, Yugo Osaka, Takuya Tsujiguchi, Akio Kodama. (2022) Dehumidification behavior of an aluminophosphate zeolite coated crossflow heat exchanger driven with direct hot water heating and evaporative cooling, *Applied Thermal Engineering* 210, no. 118355. *https://doi.org/10.1016/j.applthermaleng.2022.118355*

19. Yuanyuan Zhou, Zhen Yan, Qiumin Dai, Yanshun Yu. (2021) Experimental and numerical evaluation of a two-stage indirect/thermoelectric assisted direct evaporative cooling system, *Energy Conversion and Management* 248, no. 114790. *https://doi.org/10.1016/j.enconman.2021.114780*

20. S.A. Nada, H.F. Elattar, M.A. Mahmoud, A. Fouda. (2020) Performance enhancement and heat and mass transfer characteristics of direct evaporative building free cooling using corrugated cellulose papers, *Energy* 211, no. 118678. *https://doi.org/10.1016/j.energy.2020.118678*

21. Ibrahim Dincer, Azzam Abu-Rayash. (2020) Chapter 6 - Sustainability Modeling, *in Energy Sustainability*, Academic Press, pp. 119-164.

22. Aleksejs Prozuments, Arturs Brahmanis, Armands Mucenieks, Vladislavs Jacnevs, Deniss Zajecs. (2022) Preliminary Study of Various Cross-Sectional Metal Sheet Shapes in Adiabatic Evaporative Cooling Pads, *Energies* 15, no. 11. *https://doi.org/10.3390/en15113875*

23. U.S. Department of Energy, *Evaporative Cooler Sizing*, in EnergyPlus™ Version 22.1.0 Documentation, pp. 584-585. (Accessed Nov 2022)

24. ASHRAE, ANSI/ASHRAE Standard 133-2015 *Method of Testing Direct Evaporative Air Coolers*, Atlanta, GA: ASHRAE, 2015.

25. Department of Energy, *Engineering Reference - EnergyPlus™ Version 22.1.0 Documentation*, 29 March 2022. [Online]. *https://energyplus.net/documentation*. [Accessed Dec 2022].

26. Department of Energy, *EnergyPlus*, 2022. *https://energyplus.net/*. [Accessed on December 2022].

27. Imtiyaz Hussain, Farzana Bibi, Showkat Ahmad Bhat, Uzair Sajjad, Muhammad Sultan, Hafiz Muhammad Ali, Muhammad Waheed Azam, Sachin Kumar Kaushal, Sajid Hussain, Wei-Mon Yan. (2022) Evaluating the parameters affecting the direct and indirect evaporative cooling systems, *Engineering Analysis with Boundary Elements* 145, 211-223. *https://doi.org/10.1016/j.enganabound.2022.09.016*

28. Mingxuan Yan, Suoying He, Na Li, Xiang Huang, Ming Gao, Mengfei Xu, Jiayu Miao, Yuanshen Lu, Kamel Hooman, Jianjun Che, Zhe Geng, Shuzhen Zhang. (2021) Experimental investigation on a novel arrangement of wet medium for evaporative cooling of air, *Internal Journal of Refrigeration* 124, 64-74. *https://doi.org/10.1016/j.ijrefrig.2020.12.01*

29. Xiangjie Chen, Yuehong Su, Devrim Aydin, Xingxing Zhang, Yate Ding, David Reay, Richard Law, Saffa Riffat. (2017) Experimental investigations of polymer hollow fibre integrated evaporative cooling system with the fibre bundles in a spindle shape, *Energy and Buildings* 154, 166-174. *https://doi.org/10.1016/j.enbuild.2017.08.068*

30. M. Lata, D.K. Gupta. (2021) Experimental Investigation and Simulation of Modified Evaporative Cooling System, in *Advances in Mechanical Engineering*, Singapore, Springer, pp. 441-448. *https://doi.org/10.1007/978*

31. *AirCare Wicking Filters. https://www.aircare.com/wicking-filters/*. (Accessed Apr 2023)

32. Weichao Yan, Xiangzhao Meng, Xin Cui, Yilin Liu, Qian Chen, Liwen Jin. (2022) Evaporative cooling performance prediction and multi-objective optimization for hollow fiber membrane module using response surface methodology, *Applied Energy* 325. *https://doi.org/10.1016/j.apenergy.2022.119855*

ABOUT STUDENT AUTHOR

Michael Wilkins is a mechanical engineering major student and will graduate in Spring of 2023. He will continue into a career in the HVAC industry as a manufacturing engineer and continue research in related fields.

PRESS SUMMARY

This study reviews current literature on direct evaporative cooling and how predictive modeling is derived. The thermodynamics properties, evaporative cooling media, and types of models discussed aim to provide insight in future research efforts as to the most common and effective practices. Consistency among research allows for advancements in evaporative cooling technologies at a much faster rate.

Semantic Interpretations of Ditransitive Constructions in English

*Marcella Jurotich**

Department of Linguistics, Carleton College, Northfield, MN

https://doi.org/10.33697/ajur.2023.087

*Student: mmaxjurot@gmail.com**
Mentor: cussery@carleton.edu

ABSTRACT
This study addresses claims made by two theories—the Alternative Projection and Verb Sensitive approaches—regarding an interpretation of possession attributed to certain ditransitive constructions. The Alternative Projection approach argues that an interpretation of possession is only available in the double object (DO) pattern expressed by English ditransitive verbs (**1a**) and is not available in the prepositional (PP) pattern (**1b**). The Verb Sensitive approach argues that this possession interpretation is either available for both the DO and PP patterns, or for neither pattern, depending on the class of ditransitive verb with which the patterns occur.
(**1a**) The salesperson gave the young farmer the grain mixture.
(**1b**) The salesperson gave the grain mixture to the young farmer.
Both approaches posit a possession interpretation of the DO pattern across all ditransitive verbs. This study tests to what degree native English speakers interpret a meaning of possession from the DO and PP patterns through an online survey with 88 participants. Ditransitive verbs from five semantic classes are analyzed to determine if the interpretation of possession varies based on use of the DO or PP pattern (Alternative Projection) or by the semantic class of the verb (Verb Sensitive). The results do not support the Alternative Projection approach. The results suggest partial support for the Verb Sensitive approach, as semantic classes do not entirely follow the pattern predicted by this approach. Further, judgements reported in this study contradict some judgements reported in the literature, highlighting the importance of quantitative studies in evaluating theoretical claims.

KEYWORDS
Ditransitives in English; Ditransitive Verbs; Survey; the Dative Alternation; Semantics; Verb Semantics; Alternative Projection approach; Verb Sensitive approach

INTRODUCTION

Ditransitives are a type of verb which allow two complements. In English, many, though not all, ditransitive verbs exhibit a behavior known as the dative alternation in which the order of the complements following the verb may alternate.[1–3] The two complements in ditransitive constructions have different functions; there is the *theme* and the intended *recipient* of that theme.[3] In (**1a**), the recipient (*the young gymnast*) precedes the theme (*some practice equipment*), and both complements are expressed as noun phrases. This contrasts the reversed order of the complements in (**1b**), where the recipient, now expressed through a prepositional phrase headed by the preposition *to*, follows the theme, which is still a noun phrase.

(**1a**) The retiring athlete sold the young gymnast some practice equipment[A].
(**1b**) The retiring athlete sold some practice equipment to the young gymnast.

The ditransitive construction represented in (**1a**) is known as the double object (DO) pattern and the construction represented in (**1b**) is the prepositional (PP) pattern. Verbs that allow both the pattern in (**1a**) and (**1b**) exhibit the dative alternation. This terminology is used for English, even though there is no morphological dative case in English.

While ditransitive verbs are attested cross-linguistically,[4] the dative alternation is less widespread.[5] In a study conducted by Siewierska (1998), only around 6% of the 219 languages sampled had the dative alternation. The exact expression and scope of the dative alternation is contingent upon language-specific features.[4] For instance, some languages have only a small number of ditransitive verbs, while other languages, such as English, possess an extensive class of ditransitives.[4] In addition to language-specific features, there are a number of factors that are attested across different languages to impact both speakers' choice

[A]Uncited examples are experimental items from this study.

between the two patterns as well as the acceptability[B] of the two patterns.[6–8] Key factors influencing the dative alternation include characteristics of the complements such as their length and complexity, whether they are animate or not, and if they are new information to the speaker.[6, 9] These factors may result in (1) an otherwise acceptable pattern becoming unacceptable when those factors are in effect and vice versa or (2) a seeming preference for one pattern over the other, even if both patterns are still acceptable. For instance, in cases such as (1) where both the DO and PP patterns are available, speakers still choose between which of the two patterns to use. What motivates speakers' choice of the DO or PP pattern? This is an area of extensive but inconclusive research. The *Introduction* provides a brief overview of the literature on key cross-linguistic factors that influence the availability as well as the usage of the DO and PP patterns, before a second, related question is discussed. Do the different patterns have different meanings? This question is the target of this study.

A set of correlated features of the complements—animacy, definiteness, pronominality, and given vs. new information—affect the order of the recipient and the theme relative to each other; and thus, which pattern speakers use.[6, 10] If one complement is more animate, definite, pronominal, and/or given (information already known to the speakers) than the other, it tends to precede it in the sentence.[11] In ditransitive constructions more broadly, recipients tend to be animate, if not human, and given information, while themes are frequently inanimate, new information.[12, 13] This behavior favors the word order of the DO pattern. Indeed, in a corpus study by Bresnan et al. (2007), DO constructions were found to be eight times more likely than PP constructions to have given information before new information. Correspondingly, Bresnan (2007) reports that DO constructions with pronominal recipients occur at a higher frequency than those with full noun phrases (e.g. *a can of beer*, *John*) in corpus data. Studies on corpus data further suggest that speaker judgements do not always align with judgements reported in the theoretical literature, and that theoretical literature may not fully capture the range of constructions acceptable to speakers in more naturalistic environments.[10]

The animacy of the recipient has another effect on the acceptability of the DO pattern, as shown in (**2**), with the ditransitive verb *send*. In the PP pattern, the use of both the animate recipient *Sue* (**2a**) and the inanimate recipient *Philadelphia* (**2b**) results in acceptable sentences. However, the DO pattern is not acceptable with the inanimate *Philadelphia* (**2d**). The only interpretation in which (**2d**) becomes licit is one where *Philadelphia* is represents an organization or some other group of people—that is, an interpretation where *Philadelphia* is actually animate.[14]

(**2a**) The editor sent the article to Sue. (Harley, 2002: 7)
(**2b**) The editor sent the article to Philadelphia. (Harley, 2002: 7)
(**2c**) The editor sent Sue the article. (Harley, 2002: 7)
(**2d**) ?The editor sent Philadelphia the article. (Harley, 2002: 7)[C]

Harley (2002) attributes the unacceptability of an inanimate recipient in the DO construction to a difference in meaning between the two patterns, wherein the DO pattern has a meaning of possession, and the PP pattern has a meaning of motion. Both animate and inanimate recipients are compatible with the meaning of motion in the PP pattern; however, successful transfer of possession, as argued to be encoded by the DO pattern, requires an animate recipient. That is, the physical entity of the city of Philadelphia is incapable of possessing *the article*. Harley (2002) attests this distinction between the DO pattern introducing a meaning of possession and the PP pattern introducing a meaning of motion across all ditransitive verbs that allow the dative alternation. This approach is known as the Alternative Projection approach.[14–16]

A paper by Rappaport Hovav and Levin (2008), however, notes different behavior in ditransitive constructions with regards to the meaning of possession. For instance, Rappaport Hovav and Levin (2008) and Beavers (2011) point out the following pattern of behavior for the verb *give* and for other verbs within the same semantic class of ditransitives (*hand, lend, loan,* etc.): when both the DO pattern (**3a**) and PP pattern (**3b**) are followed by a clause denying a successful transfer of possession of the theme to the recipient, the sentence is less acceptable. This suggests that, at least for this set of ditransitive verbs, ditransitive constructions with both the DO and PP pattern have a meaning of possession. This approach is known as the Verb Sensitive approach.[2–3]

(**3a**) #John gave/handed Mary the salt, but he dropped it before she got it. (Beavers, 2011: 10)
(**3b**) #My aunt gave/lent/loaned some money to my brother for new skis, but he never got it. (Rappaport Hovav and Levin, 2008: 37)

[B]Acceptability is a distinct metric from grammaticality. Acceptability differs from grammaticality in that it is a measure of speakers' performance of a language and is directly testable. Speakers' judgements on the acceptability of items inform grammaticality, which concerns the abstract rules that make up a language. This study tests the *acceptability* of ditransitive constructions but not the grammaticality.
[C]The (?) and (#) in front of an item indicate respectively that the item is not fully licit or is ill-formed. An (*) in front of an item indicates a fully unacceptable item that violates the rules of a language's grammar.

Notably, the unacceptability of (**2d**) and (**3**) occurs when the givenness and complexity of the complements are controlled for, indicating that variations in those features cannot account for this unacceptability.

Both the Alternative Projection and Verb Sensitive approaches agree that certain ditransitive constructions have a meaning of possession; however, they diverge in regard to the source of the possession meaning. Namely, the Alternative Projection approach argues that a possession interpretation stems from the DO pattern and is thus available only in this pattern, whereas the Verb Sensitive approach argues that the possession interpretation stems from the verb and is such either available for both the DO and PP patterns or available for neither pattern depending on the ditransitive verb that the patterns occur with. That is, according to each theory respectively, either the DO and PP constructions pattern differently regarding a meaning of possession, or different classes of ditransitive verbs pattern differently in regard to a meaning of possession.

There is a difference as well in the scope and means of argumentation between the Alternative Projection and Verb Sensitive approaches. The Alternative Projection approach posits that the difference in meaning between the DO and PP patterns is determined by their different structures.[16] This is a primarily syntactic distinction, and studies under the Alternative Projection approach often concentrate on determining the syntactic structure of the DO and PP constructions.[14–16] As the Verb Sensitive approach attributes the entailment of possession to the core meaning of the verb, it more so focuses on the semantics of ditransitive constructions.[2–3] The orientation of both of these studies, however, is primarily theoretical in nature, and as such, neither of these studies provides quantitative results as to the entailment of possession in patterns or verb classes.

Indeed, many approaches to verb meaning in linguistics posit a distinction between the core meaning of the verb, which applies to all uses of that verb regardless of context, and the meaning of the construction representing the event, in this specific case, the DO and PP patterns.[2, 17] Under both the Alternative Projection and Verb Sensitive approaches, the inherent meaning of the DO construction by itself is possession. However, in the Verb Sensitive approach, the possession encoded by the DO pattern is actually *prospective* possession, meaning that possession of the theme by the recipient only needs to obtain in some possible worlds, not necessarily the actual world. Rather, in the Verb Sensitive approach, certain verbs, whose core meaning entails actual possession, subsume the independent meaning of the DO and PP pattern, such that when the DO and PP patterns occur with this given verb, both patterns have an interpretation of actual, obtained possession.[3] These two approaches further diverge with respect to the meaning of the PP pattern. While the Alternative Projection approach argues that the PP pattern has a meaning of motion, the Verb Sensitive approach expands on the potential meanings of the PP pattern to be motion or possession. The availability of these interpretations depends on which class of ditransitive verb the PP pattern occurs with. Notably, this meaning of possession in the PP pattern is only ever prospective; under the Verb Sensitive approach, actual entailment of possession, such that its denial is infelicitous with the meaning of the entire ditransitive construction, stems only from the verb itself. Further, the Alternative Projection approach is not uniform in its claims on the meaning of the two patterns. While Harley (2002) and Krifka (1999), among others, attest that the DO construction entails actual possession,[17] Harley and Jung (2015), a later paper under the Alternative Projection approach, adopts the proposal from Beavers (2011) that the possession meaning of the DO pattern is solely prospective, and thus can be denied. This study focuses on the claims made by earlier papers under the Alternative Projection Approach, such as Harley (2002) and Krifka (1999), although there is further discussion of Harley and Jung (2015) in the *Discussion*.

Building on existing English verb classification schemes developed by Pinker (1989) and Levin (1993), the Verb Sensitive approach groups ditransitive verbs into semantic classes based on the general meaning they have. This study focuses on five semantic classes drawn from Rappaport Hovav and Levin (2008) and Beavers (2011), both papers arguing in support of the Verb Sensitive Approach: Giving, Future Having, Communication, Instrument of Communication, and Motion. The meaning of the DO and PP patterns for a given verb depends on which of these five classes the verb belongs to. For verbs in the Giving class only, the inherent possession meaning of the verb subsumes the independent meaning of the DO and PP constructions. Although the exact meaning of a verb varies, verbs in the Giving category all inherently signify acts of giving. None of the remaining classes of verbs inherently entail a meaning of successful possession.[2, 3] Verbs in the Future Having class signify acts of *future* giving or possession, and verbs in the Communication class are verbs of communicated message. Instrument of Communication verbs have a meaning of communicated message through a specific instrument (a phone, fax machine, *etc.*) and Motion verbs have a meaning of sending motion or an instantaneous causation of ballistic motion.

This study targets arguments made by the Alternative Projection and Verb Sensitive approaches that in certain ditransitive constructions a meaning of possession is entailed. Specifically, this study tests whether the DO pattern entails possession while the PP pattern does not or whether Giving verbs entail possession for both patterns while other classes do not. The aims are twofold. First, judgements reported in the literature by Rappaport Hovav and Levin (2008) and Beavers (2011) are checked through experimental methodology to see if they are corroborated by speakers. The second goal is to ascertain which theoretical approach best accounts for variation in pattern in English ditransitives. Participants went through a two-step process of rating experimental items in an online survey. First, they rated a ditransitive construction with either the DO or PP pattern. Then, they were prompted with a clause that denies any potential transfer of possession that could be entailed in the ditransitive construction

(either through the verb or the DO pattern). Speaker judgements on the ditransitive constructions from the first set of experimental items are predicted to conform to attested acceptability judgements from the literature**.** However, participants' judgements on the acceptability of the second set of experimental items is the target of the core research question. If the data are in line with the Alternative Projection approach, and the DO pattern entails possession while the PP pattern does not, the behavior detailed in (**4**) is expected. If the Verb Sensitive approach holds true, however, the data should exhibit all of the three behaviors described in (**5**). As the Verb Sensitive approach argues for an analysis of ditransitives based on the class of ditransitive verb, analyzing ditransitives by their category rather than solely as individual verbs allows us to test the Verb Sensitive approach in addition to the Alternative Projection approach.

(**4**) Denial of Possession will not be possible following the DO pattern, since the DO pattern entails possession, and the PP pattern has a meaning of motion.
(**5a**) Denial of possession will not be possible after both the DO and PP patterns in the Giving class.
(**5b**) Denial of possession will be possible following both the DO and PP patterns in the Future Having, Communication, Instrument of Communication, and Motion classes.
(**5c**) Denial of possession after both the DO and PP patterns will be rated significantly worse in the Giving class than in the Future Having, Communication, Instrument of Communication, and Motion classes.

Notably, for there to be sufficient evidence to fully support the Verb Sensitive approach, all three predictions must be attested in the data.

METHODS AND PROCEDURES

Eight-eight participants between the ages of 18 and 35 took part in an online survey administered through Qualtrics. All participants were native English speakers and had not taken linguistics classes. Upon completion of the survey, participants were compensated with a $20 gift card. The experimental design consisted of two steps: an initial set of ditransitive sentences and a follow-up set of denial of possession clauses. Participants were first presented with a ditransitive clause, containing either a DO construction or a PP construction. Participants rated the acceptability of this initial sentence on a 1-7 Likert scale. If participants rated the sentence as a 4 or higher, they were prompted with a follow-up clause denying any potential transfer of possession entailed in the initial ditransitive sentence. The follow-up clause was not ditransitive. The overall structure of the initial item and the follow-up item (underlined) is shown for the DO pattern (**6a**) and PP pattern (**6b**).

(**6a**) Nicolas handed his baby sister a giant lollipop… but their dad snatched it away before she could grab it, and the little girl looked longingly at the lollipop.
(**6b**) Nicolas handed a giant lollipop to his baby sister… but their dad snatched it away before she could grab it, and the little girl looked longingly at the lollipop.

In order to ensure that the participants rated the denial of possession clause as a continuation of the initial ditransitive construction, the initial item was presented along with the follow-up item in the prompt, and participants were specifically instructed to rate how natural they found the follow-up clause as a continuation of the initial item. Participants were also given practice items at the beginning of the survey. The online survey consisted of 30 target items, counting the initial ditransitive sentence and its follow-up as a single unit. As the Verb Sensitive approach focuses on the behavior of classes of ditransitives, 30 ditransitive verbs evenly distributed across five semantic classes were tested, adapted from Rappaport Hovav and Levin (2008) and Beavers (2011). These classes are Giving, Future Having, Communication, Instrument of Communication, and Motion (**Table 1**). Further, the experimental items were constructed in a similar format to examples from Rappaport Hovav and Levin (2008) and Beavers (2011), so as to test if those judgements are corroborated by speakers (see (**3**)).

Giving (G) - *Hand, Give, Lend, Sell, Rent, Pay*
Future Having (FH) - *Offer, Grant, Owe, Promise, Allocate, Bequeath*
Communication (COM) - *Show, Write, Teach, Tell, Explain, Announce*
Instrument of Communication (IC) - *Email, Email.2, Text, Text.2, Telegraph, Fax*
Motion (M) – *Send, Send.2, Forward, Kick, Toss, Throw*

Table 1. Semantic classes of ditransitive verbs based on Rappaport Hovav and Levin (2008) and Beavers (2011).

Each ditransitive verb in the survey had two separate entries: one with the DO pattern and one with the PP pattern. For a given ditransitive verb, participants were presented with either the DO pattern or the PP pattern. That is, no single participant was presented with both patterns for the same verb. In total, participants encountered 15 DO constructions and 15 PP constructions. The order of the experimental items was randomized for each participant. This survey tested the relative acceptability of each pattern. As a given participant only viewed one pattern per verb, data on the preference of one pattern over the other could not be gathered. Animacy and pronominality, known factors impacting the usage and acceptability of both patterns, were controlled for.[6, 9, 18] All themes were inanimate, all recipients were animate, and no recipient or theme was a pronoun. All ditransitive verbs in

the survey were conjugated in the simple past. In addition to the 30 target items and their corresponding denial of possession clauses, 50 filler items were interspersed throughout the survey. These filler items consisted of various non-ditransitive sentences whose judgements by native speakers are consistent and clearly established. They were a mix of grammatical or ungrammatical as well as plausible or implausible sentences. The filler items served as a check to determine if a participant was defaulting to a certain rating throughout the survey or offering judgements without fully reading the item, and to accordingly allow the removal of that participant's responses from the analysis. A one-way ANOVA was run through Python; p-values below 0.05 were considered statistically significant.

This study is part of a larger study investigating Icelandic and Faroese ditransitives, with English ditransitives used for comparison, and experimental items were the same across all three languages. Acceptability judgements collected on the initial ditransitive sentences serve two key purposes: (1) to prevent participants who disliked a ditransitive construction from rating the follow-up, and (2) to gather parallel information on all three languages. As Icelandic and Faroese ditransitives are understudied, information on acceptability judgements is particularly crucial. However, the data from English acceptability judgements also yield intriguing findings contributing to the literature. To facilitate this parallel methodology, the verbs *send* (COM), *text* (IC), and *email* (IC) in the English survey had two entries. This is because there are not enough unique verbs in these verb classes in Icelandic and Faroese, so multiple entries for individual verbs were created to ensure an equal number of experimental items in each semantic class. Due to the understudied nature of Icelandic and Faroese ditransitives, further research calibrated to those languages has been conducted.

RESULTS

Initial Ditransitive Sentences

The results support the literature which attests that the dative alternation occurs in this set of ditransitive verbs. Two semantic classes exhibited a significant difference in ratings between the DO pattern and the PP pattern for the initial ditransitive sentences. Participants rated the DO pattern significantly lower than the PP pattern with $p < 0.001$ in both the Communication and the Motion classes (**Figure 1**). The PP pattern was rated significantly higher for six individual verbs: *tell*, *explain*, and *announce* in the Communication class; *kick* and *toss* in the Motion class; *fax* in the Instrument of Communication class. The DO pattern was rated significantly higher for one verb, *show*, in the Communication class. No other verbs in the Communication and Motion classes exhibited a significant difference between patterns.

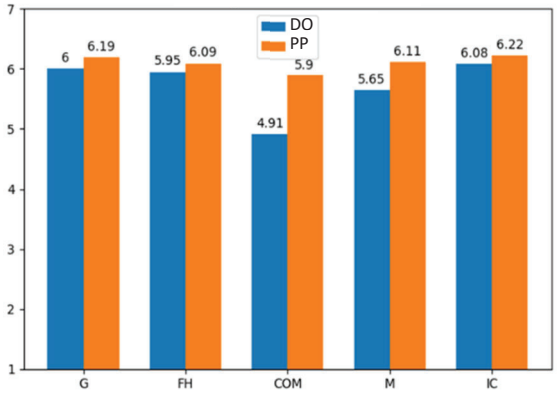


Figure 1. Average ratings of DO and PP patterns for each semantic class.

Follow-up Denial of Possession Clauses

Data from the follow-up denial of possession clauses do not support the Alternative Projection approach and provide partial support for the Verb Sensitive approach. Denial of possession after the DO construction was not consistently rated significantly lower across all the verbs tested. Participants rated the denial of possession clause following the PP pattern significantly lower in the Motion class ($p < 0.01$) (**Figure 2**). No other semantic classes exhibited a significant difference between the denial of possession after the DO and PP patterns. There were significant differences between follow-up ratings after the DO and PP patterns for some individual verbs. Denial of possession after the DO pattern was rated significantly worse for *hand* in the Giving class (**Figure 3**) and for *tell* in the Communication class (**Figure 5**). Denial of possession after the PP pattern was rated significantly worse for *owe* and *allocate* in the Future Having class (**Figure 4**). Denial of possession after the PP pattern was rated significantly worse for two verbs, *kick* and *throw*, in the Motion class (**Figure 6**). There was no significant difference in follow-up ratings between patterns for Instrument of Communication verbs as a class or for individual verbs within the class (**Figure 7**).

The average follow-up ratings after both patterns were lower than the average ratings for the initial experimental items, except for the DO pattern in the Communication class (**Figure 1, Figure 2**).

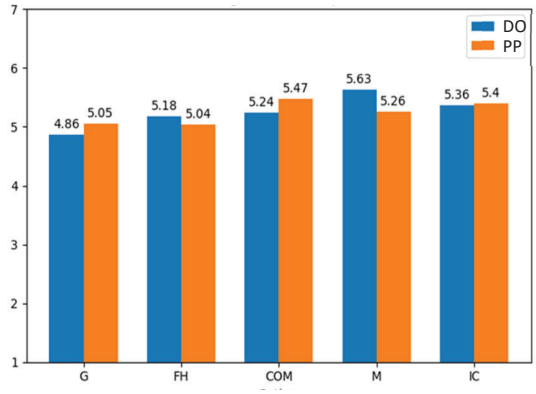


Figure 2. Average ratings of denial of possession clauses following DO and PP patterns for each semantic class.

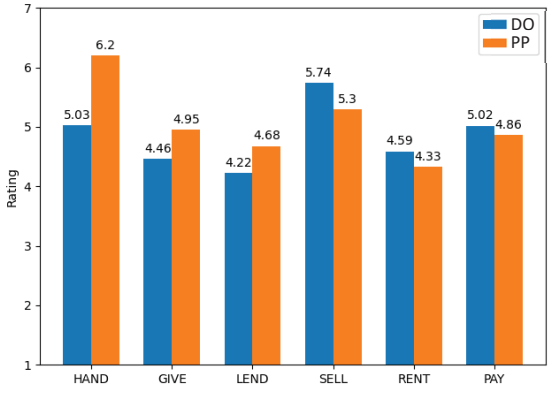


Figure 3. Average ratings of denial of possession clauses following DO and PP patterns for Giving class.

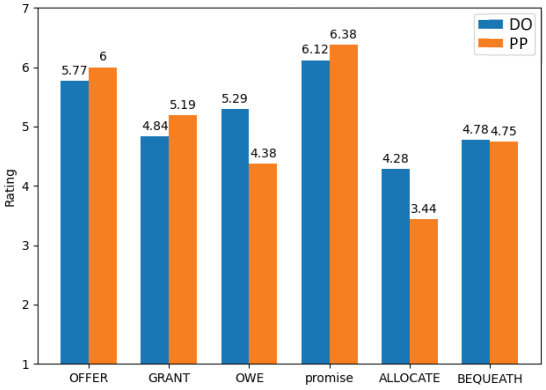


Figure 4. Average ratings of denial of possession clauses following DO and PP patterns for Future Having class.

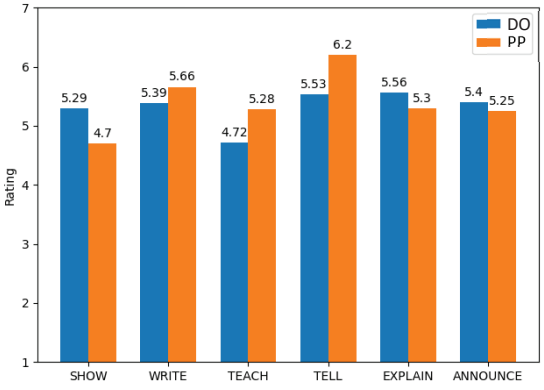


Figure 5. Average ratings of denial of possession clauses following DO and PP patterns for Communication class.

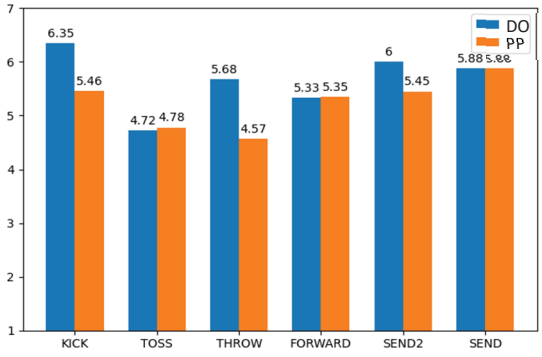


Figure 6. Average ratings of denial of possession clauses following DO and PP patterns for Motion class.

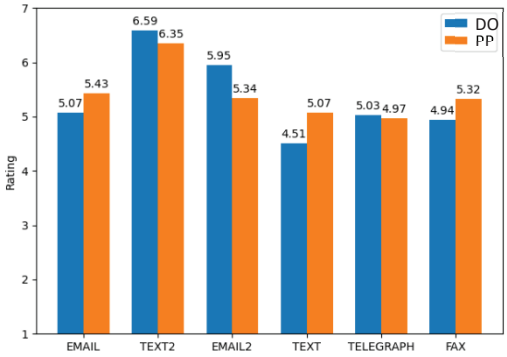


Figure 7. Average ratings of denial of possession clauses following DO and PP patterns for Instrument of Communication class.

DISCUSSION

Initial Ditransitive Sentences

The results do not indicate preference for either pattern, as an individual participant only encounters either the DO or PP pattern for a given ditransitive, but rather the relative acceptability of each pattern. On the whole, the results corroborate the literature which attests that the dative alternation occurs in this set of ditransitive verbs.[2] For two verbs in the Communication class, *explain* and *announce*, the PP pattern is rated significantly higher than the DO pattern ($p < 0.001$). *Explain* and *announce* do not undergo the dative alternation; the inclusion of these verbs in the survey served as an internal check to assess the reliability of participant

responses. Although two out of the five semantic classes displayed a significant difference in ratings between patterns, only four individual verbs within those categories exhibited a significant difference. These verbs were concentrated in the Motion class (*kick* and *toss*) and Communication class (*tell* and *show*); *fax* (IC) was the only verb with a significant difference in ratings between patterns whose overall class did not also have a significant difference (**7**) (average ratings in brackets).

(**7a**) The whistleblower faxed the government officials the confidential documents. [5.11]
(**7b**) The whistleblower faxed the confidential documents to the government officials. [5.93]

For only one verb *show* (COM), participants rated the DO pattern significantly higher than the PP pattern. No classes rated the DO pattern significantly higher. However, the ratings for the majority of the semantic classes and individual verbs examined do not significantly differ by pattern and the average ratings for both patterns are acceptable (4 and higher) with the exception of *kick* and *show*. The average rating of the DO pattern for *kick* was 3.64 and the average rating of the PP pattern for *show* was 4.11, both of which were lower than expected in verbs that exhibit the dative alternation.

This overall trend towards higher relative acceptability of the PP pattern is surprising, particularly in light of Stefanowitsch and Gries's (2003) claim of a strong association with the DO pattern for verbs such as *give*, *tell*, and *send* among others. Stefanowitsch and Gries (2003) use corpus data to calculate the frequency with which a given ditransitive verb co-occurs with a DO construction, and thereby the strength of the association between a ditransitive verb and the DO construction. Specifically, their methodology looked for the frequency of a verb in a DO construction in relation to frequency of that verb in all other constructions and frequency of all other verbs in the DO construction. Stefanowitsch and Gries (2003) do not explicitly test the association between these same verbs and the PP pattern; however, the overall strength of association with the DO found in their results does suggest that there should be a higher rating of the DO pattern than the PP pattern. However, in the results from this current study, the PP pattern is rated significantly higher than the DO pattern for *tell* ($p = 0.007$). While there is no significant difference for *give* and both experimental items for *send*, the average ratings for the PP patterns are higher for all three items. Despite the generally higher acceptability of the PP pattern in the data from this study, only a small percentage of verbs exhibit any significant difference between patterns.

Follow-up Denial of Possession Clauses
Ratings for the follow-up items vary extensively across the semantic class of ditransitive and by individual verbs within each class. This variation is not in line with the predictions of this study, as outlined in (**4**) and (**5**). The DO construction is not consistently rated significantly lower than the PP construction after denial of possession, and the DO and PP constructions do not pattern the same way across and within verb classes with respect to denial of possession. These patterns offer insight into which approach best fits the data.

Under the Alternative Projection approach, difference in structure results in difference in meaning, and the meanings of the DO and PP patterns generally do not overlap.[16] Only in one class, Motion, is there a significant difference between patterns with denial of possession rated significantly lower after the PP pattern ($p < 0.01$), suggesting that the PP construction was more strongly associated with possession than the DO construction. Thus, the results do not corroborate predictions from the Alternative Projection approach (see (**4**)).

According to the Verb Sensitive approach, for no class of ditransitive verb, nor individual verb, should there be any significant difference between the DO and PP constructions in terms of the acceptability, or lack thereof, of denial of possession. Verbs that entail possession, entail it for both constructions, and verbs that do not, entail possession for neither construction.[2–3] However, as the Verb Sensitive approach posits that different classes of ditransitive verbs pattern differently with respect to the entailment of possession, there should be a significant difference in acceptability of denial of possession between certain classes of verbs. Specifically, should the Verb Sensitive approach hold, the following patterns are expected (**5**).

(**5a**) Denial of possession will not be possible after both the DO and PP patterns in the Giving class.
(**5b**) Denial of possession will be possible following both the DO and PP patterns in the Future Having, Communication, Instrument of Communication, and Motion classes.
(**5c**) Denial of possession after both the DO and PP patterns will be rated significantly worse in the Giving class than in the Future Having, Communication, Instrument of Communication, and Motion classes.

The data from this study partially supports (**5b**). Denial of possession is possible after both the DO and PP constructions in the Future Having, Communication, Instrument of Communication, and Motion classes; however, the significantly lower ratings after the PP pattern in the Motion class are not in line with the predictions of the Verb Sensitive approach in (**5b**). This suggests that the PP pattern entails possession while the DO pattern does not, a behavior attested by neither the Alternative Projection nor the Verb Sensitive approaches. However, the data do not support the predictions in (**5a**) and (**5c**). Both (**5a**) and (**5c**) concern the behavior of the Giving class. Namely, both hypotheses target the claim that all verbs in the Giving class entail successful transfer

of possession, such that its denial is infelicitous with a key meaning component of the entire ditransitive construction, and that for this reason, any denial of that transfer is unacceptable.[2–3] Consequently, denial of possession should not be possible within this class (**5a**) and should be rated significantly lower after both patterns in the Giving class than in the other classes (**5c**). This is not supported by the data. The average ratings for denial of possession after the DO and PP patterns for the Giving class were 4.86 and 5.05 respectively (**Figure 2**), which do not meet the threshold for unacceptability of a rating of under 4. Further, denial of possession after both the DO and PP patterns was not rated significantly worse in the Giving class than in the Future Having, Communication, Instrument of Communication, and Motion classes (**Figure 2**).

The behavior of the Giving and Motion classes is unattested by either the Alternative Projection or Verb Sensitive approaches, and as such, merits a closer examination. While the behavior of both classes as a whole is surprising, individual verbs within these categories also deviate from the predictions of the Verb Sensitive approach[D]. In particular, denial of possession is rated significantly worse after the DO construction than after the PP construction for *hand* in the Giving class ($p < 0.01$). Not only is this not in line with the expected behavior of Giving verbs under the Verb Sensitive approach, but this explicitly contradicts examples from Beavers (2011). Namely, Beavers (2011) notes denial of possession to be unacceptable after both the DO (**8a**) and PP (**8b**) pattern for *hand*. Yet, in the results from this study, average ratings for both patterns remain high, particularly in the PP pattern (**9**; average ratings in brackets). Indeed, the average rating for the follow-up after the PP construction with *hand* is 6.2, which is markedly higher than the average ratings for either pattern for the other Giving verbs, and indeed, quite high for all verbs tested (**Figure 3**). This indicates that (1) the DO pattern seems to have more of an interpretation of possession than the PP pattern, and (2) there is a relatively high acceptability of denial of possession after the PP pattern.

(**8a**) #John gave/handed Mary the salt, but he dropped it before she got it. (Beavers, 2011: 10)
(**8b**) #Carla handed the money to Norm, but he never received it. (Beavers, 2011: 43)
(**9a**) Nicolas handed his baby sister a giant lollipop… but their dad snatched it away before she could grab it, and the little girl looked longingly at the lollipop. [5.03]
(**9b**) Nicolas handed a giant lollipop to his baby sister… but their dad snatched it away before she could grab it, and the little girl looked longingly at the lollipop. [6.2]

These judgements for *hand* contrast to the judgements for the rest of the verbs in the Giving category. There is some minor variation in the average denial of possession ratings for the remaining Giving verbs, with *sell* having slightly higher acceptance of denied possession than *rent*, which is on the lower end of denial of possession ratings for this class (**10**). Overall, with the exception of *hand*, Giving verbs have lower, but not unacceptable ratings, for denial of possession after both patterns.

(**10a**) The clerk at the help desk rented the American tourists a car… but their credit card was denied so that they had to leave without the rental car. [4.59]
(**10b**) The clerk at the help desk rented a car to the American tourists… but their credit card was denied so that they had to leave without the rental car. [4.33]

Similar verb-by-verb variation is attested in the results for the Motion class. While denial of possession after the PP pattern was rated significantly worse for two verbs, *kick* and *throw*, there was remarkably minimal variation between the average denial of possession ratings between the DO and PP constructions for the other Motion verbs (**Figure 6**). While the results for *kick* are potentially complicated by the low average rating of the initial DO construction at 3.64, only participants that judged an initial ditransitive construction as acceptable proceeded to rate the denial of possession follow-up. Further, Rappaport Hovav and Levin (2008) attest denial of possession after both patterns for *throw* as acceptable (**11**), since *throw* as a verb in the Motion class does not entail possession. However, the results of this study contradict this and indicate that denial of possession after the PP pattern is significantly worse than after the DO pattern (**12a**), and the average ratings for both constructions in (**12b**) are towards the lower end of the rating data.

(**11**) I threw the ball to Julian, but it fell short of him. (Rappaport Hovav and Levin, 2008: 34)
(**12a**) The excited little girl threw the cute dog a bouncy ball… but the ball ended up on the roof, so the dog never got his ball. [5.68]
(**12b**) The excited little girl threw the bouncy ball to the cute dog… but the ball ended up on the roof, so the dog never got his ball. [4.57]

Variation in the behavior of individual verbs is not unique to the Giving and Motion classes. Indeed, as mentioned in the *Results*, a number of verbs (*owe* and *allocate* in Future Having and *tell* in Instrument of Communication) also exhibit significant differences

[D] This study was designed to evaluate the data based on overall patterns of constructions or verb classes, and as such, there is only a single entry per verb, with the exceptions of *send*, *text*, and *email*. Thus, any discussion of individual verbs, rather than overall semantic classes, is inherently limited.

between denial of possession judgements after the two patterns. Deviation of individual verbs from the predictions of both approaches is unexpected. This variation cannot be accounted for under the Alternative Projection approach, as the entailment of possession stems from the DO construction and is as such is invariant and independent of the verb. Another possible account is that, under the Verb Sensitive approach, entailment of possession is sensitive to individual verbs more so than, or in addition to, overall classes. Yet, if this were the case, even in verbs that varied in the meaning they entail, there would be no significant difference between the DO and PP patterns with regards to possession interpretations. Finally, both Harley and Jung (2015), in a revised Alternative Projection approach, along with Beavers (2011) and Rappaport Hovav and Levin (2008) argue that the DO pattern has a meaning of prospective possession. The behavior of *hand* and *tell*, wherein denial of possession is rated significantly worse after the DO construction, is incongruent with these claims.

Certainly, quantitative results from this study indicate that participant judgements on individual experimental items contradict claims made by Rappaport Hovav and Levin (2008) and Beavers (2011) regarding their acceptability. Despite these differences, participants' ratings of the follow-up clauses do correspond to some judgements reported in Rappaport Hovav and Levin (2008) and Beavers (2011). Denial of possession does not result in unacceptability for *send* regardless of the pattern both in Rappaport Hovav and Levin (**13**) and this study (**14**).

(**13a**) Lewis sent/shipped Sam a bicycle, but it never arrived. (Rappaport Hovav and Levin, 2008: 42)
(**13b**) Lewis sent/shipped a bicycle to Sam, but it never arrived. (Rappaport Hovav and Levin, 2008: 42)
(**14a**) The thoughtful managers sent the hardworking secretary a surprise birthday present… but the mailman misplaced it, and the secretary never got her gift. [5.88]
(**14b**) The thoughtful managers sent a surprise birthday present to the hardworking secretary… but someone stole the package from the mailroom and the secretary never got her gift. [5.88]

Indeed, as Bresnan (2007) notes, mismatches between judgements reported in the literature and speaker judgements are not uncommon. Acceptability judgements obtained in more naturalistic environments may capture a broader range of acceptable constructions than reported in the literature.[10]

CONCLUSIONS

Extensive theoretical research alongside emerging experimental work attempts to account for the phenomenon of the dative alternation in English. This research is, however, inconclusive, and open questions remain concerning how features of the complements, such as animacy and givenness, and variation in meaning between patterns affect the use and availability of each pattern.[2, 3, 7, 10, 16, 19, 20] This study targets the claim that there is a difference in meaning between either the DO and PP patterns or between ditransitive constructions in certain classes of ditransitive verbs. Specifically, this study tests whether the DO pattern entails possession and the PP pattern does not (in line with the Alternative Projection approach) or whether the entailment of possession depends on the overall semantic class of a verb (in line with the Verb Sensitive approach). These two accounts are theoretical in nature; this study uses experimental methodology to test if speaker judgements align with the literature.

The results do not support the Alternative Projection approach but suggest partial support for the Verb Sensitive approach. Denial of possession is not rated consistently significantly worse after the DO construction. Denial of possession is possible after both the DO and PP constructions in the Future Having, Communication, Instrument of Communication, and Motion classes, supporting the Verb Sensitive approach. However, the significantly lower ratings after the PP pattern in the Motion class do not align with the Verb Sensitive approach. Further, denial of possession was in fact, on average, acceptable after both constructions in the Giving class, and neither were the follow-up ratings for the Giving class significantly lower than the other classes. Judgements collected in this study contradicted some judgements reported in the theoretical literature, highlighting the importance of experimental methodology in evaluating theoretical claims.

FUTURE DIRECTIONS

Adjustments to the experimental design of the study could offer greater clarity into the applicability of the Verb Sensitive approach. A reduction in the length of the survey could avoid fatigue on the part of the participants. Further, despite the benefits of collecting separate judgements on the initial ditransitive sentences and the follow-up clauses, the separation of these items is a source of potential confusion to participants. Now that the acceptability of the initial set of ditransitive verbs in both patterns is attested, streamlining the experimental design by combining the initial sentences and follow-up clauses into a single item would more directly target judgements on the denial of possession clauses. Contradictions in judgements on individual ditransitive constructions between this study and the literature underscore the value of quantitative studies. Further, as these deviations on the basis on individual ditransitives remain unaccounted for by either this study or the Alternative Projection and Verb Sensitive approaches, further quantitative studies, with a greater number of experimental items per verb, would be of merit.

ACKNOWLEDGEMENTS

First and foremost, the author thanks Dr. Cherlon Ussery for her mentoring and advice throughout the research and writing stages of this project. The author greatly thanks Anna Grove and Jack Uchitel for their extensive work on the development of experimental items, literature reviews, and analysis. Further thanks go to Helen Murphy and Alex Zhai for their contributions to the development of this project and Zach Kulstad for his work on the data analysis. This project is associated with research funded by the Icelandic Research Fund, Rannís, for the project *Ditransitives in Insular Scandinavian*, Grant #195926-051, Co-PIs, Jóhannes Gísli Jónsson and Cherlon Ussery. Support was also provided by the Humanities Center and the Dean of the College Office at Carleton College.

REFERENCES

1. Larson, R. K. (1988) On the double object construction, *Linguistic Inquiry*, 19(3), 335–391.
2. Rappaport Hovav, M.., and Levin, B. (2008) The English dative alternation: The case for verb sensitivity, *Journal of Linguistics*, 44(1), 129–167. *https://doi.org/10.1017/S0022226707004975*
3. Beavers, J. (2011). An aspectual analysis of ditransitive verbs of caused possession in English. *Journal of semantics*, *28*(1), 1-54. *https://doi.org/10.1093/jos/ffq014*
4. Haspelmath, M. (2015) Ditransitive constructions, *Annual Review of Linguistics*, 1(1), 19–41. *https://doi.org/10.1146/annurev-linguist-030514-125204*
5. Siewierska, A. (1998) Languages with and without objects: The Functional Grammar approach, *Languages in Contrast*, 1(2), 173–190. *https://doi.org/10.1075/lic.1.2.05sie*
6. Bresnan, J., Cueni, A., Nikitina, T., and Baayen, R. H. (2007) Predicting the dative alternation, in *Cognitive Foundations of Interpretation* (Boume, G., Kraemer, I., and Zwarts, J., Eds.), 69–94, Royal Netherlands Academy of Science, Amsterdam.
7. Kizach, J., and Balling, L. W. (2013) Givenness, complexity, and the Danish dative alternation, *Memory & Cognition*, 41(8), 1159–1171. *https://doi.org/10.3758/s13421-013-0336-3*
8. Köhne, J., Pickering, M. J., and Branigan, H. P. (2014) The relationship between sentence meaning and word order: Evidence from structural priming in German, *Quarterly Journal of Experimental Psychology*, 67(2), 304–318. *https://doi.org/10.1080/17470218.2013.807855*
9. Arnold, J. E., Losongco, A., Wasow, T., and Ginstrom, R. (2000) Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering, *Language* 76(1), 28–55. *https://doi.org/10.1353/lan.2000.0045*
10. Bresnan, J. (2007) Is syntactic knowledge probabilistic? Experiments with the English dative alternation, *Roots: Linguistics in Search of its Evidential Base* 2, 75–96. *https://doi.org/10.1515/9783110198621.75*
11. Brown, M., Savova, V., and Gibson, E. (2012) Syntax encodes information structure: Evidence from on-line reading comprehension, *Journal of Memory and Language*, 66(1), 194–209. *https://doi.org/10.1016/j.jml.2011.08.006*
12. Givón, T. (1979) On Understanding Grammar, Academic Press, New York.
13. Thompson, S. A. (1990) Information Flow and 'Dative Shift' in English, in *Development and Diversity: Linguistic Variation across Time and Space* (Edmondson, J., Feagin, K., and Mühlhäusler, P., Eds.), 239–253, Summer Institute of Linguistics, Dallas.
14. Harley, H. (2002) Possession and the double object construction, *Linguistic Variation Yearbook*, 2(1), 31–70. *https://doi.org/10.1075/livy.2.04har*
15. Beck, S., and Johnson, K. (2004) Double objects again, *Linguistic Inquiry* 35(1), 97–123. *https://doi.org/10.1162/002438904322793356*
16. Harley, H., and Jung, H. K. (2015) In support of the P have analysis of the double object construction, *Linguistic Inquiry*, 46(4), 703–730. *https://doi.org/10.1162/LING_a_00198*
17. Goldberg, A. E. (1995) Constructions: A Construction Grammar Approach to Argument Structure, The University of Chicago Press, Chicago.
18. Wasow, T., and Arnold, J. (2003) Post-verbal constituent ordering in English, *Topics in English Linguistics*, 43, 119–154. *https://doi.org/10.1515/9783110900019.119*
19. Bruening, B. (2010) Double object constructions disguised as prepositional datives, *Linguistic Inquiry*, 41(2), 287–305. *https://doi.org/10.1162/ling.2010.41.2.287*
20. Hallman, P. (2015) Syntactic neutralization in double object constructions, *Linguistic Inquiry*, 46(3), 389–424. *https://doi.org/10.1162/LING_a_00187*
21. Krifka, M. (1999). Manner in dative alternation. *West Coast Conference on Formal Linguistics*, 18, 260-271.
22. Pinker, S. (1989) Learnability and Cognition: The Acquisition of Argument Structure, MIT Press, Cambridge.
23. Levin, B. (1993) English Verb Classes and Alternations: A Preliminary Investigation, The University of Chicago Press, Chicago.
24. Stefanowitsch, A., and Gries, S. T. (2003) Collostructions: Investigating the interaction of words and constructions, *International Journal of Corpus Linguistics*, 8(2), 209–243. *https://doi.org/10.1075/ijcl.8.2.03ste*

ABOUT STUDENT AUTHOR

Marcella Jurotich graduated from Carleton College in June 2021 with a bachelor's degree in Linguistics and German.

PRESS SUMMARY

English has a large class of verbs, known as ditransitive verbs, that have two complements. Many, though not all of these ditransitive verbs allow for two patterns of expression of their complements (**1**).

(**1a**) The salesperson gave the young farmer the grain mixture. *Double Object (DO) pattern*

(**1b**) The salesperson gave the grain mixture to the young farmer. *Prepositional (PP) pattern*

At first glance, these patterns seem to be able to alternate freely. However, two competing theories attest that there is a difference in meaning either based on the pattern or the type of ditransitive verb that is used. Both theories argue that there is meaning of successful transfer of possession for certain ditransitive constructions—that is, that *the young farmer* successfully received the *grain mixture* given by *the salesperson*. While one theory claims that the only DO pattern has this possession meaning, the other posits that this possession meaning stems from verbs of the Giving category (*give, hand, rent,* etc.), and as such, that this meaning extends to both the DO and PP patterns when they occur with these verbs. This study evaluates these two theories through an online survey testing to what degree native speakers interpret a meaning of possession from the DO and PP patterns. The results of this study suggest that this meaning of successful possession is not restricted to the DO pattern and provide partial support for the claim that the availability of the possession meaning depends on the type of ditransitive verb. Further, judgements collected in this study contradicted some judgements reported in the theoretical literature, highlighting the importance of experimental methodology in evaluating theoretical claims.

Are Wrist-based Heart Rate Monitors a Valid Tool for Fitness Professionals to Measure Training Intensity During Exercise Classes?

Korey Little[a], John C. Sieverdes[*a], D. David Thomas[b,c], M. Blake Lineberger[a], Daniel B. Bornsteind, Marco Bergamine, & Wesley D. Dudgeon[a]*

[a]Department of Health and Human Performance, College of Charleston, Charleston SC
[b]Informatics Programs, University of Illinois, Champaign, IL
[c]Stephens Family Clinical Research Institute, Carle Health, Urbana, IL
[d]DBornsteinSolutions, LLC, Norwich, VT
[e]Department of Medicine, University of Padova, Padova, Italy

https://doi.org/10.33697/ajur.2023.088

Student email: littlekc@g.cofc.edu
Mentor email: sieverdesjc@cofc.edu, ddt3@illinois.edu, linebergerbb@cofc.edu, dan@dbornsteinsolutionsllc.com, marco.bergamin@unipd.it, dudgeonw@cofc.edu*

ABSTRACT
This article aims to inform personal trainers and group fitness coaches about the validity and utility of wrist-located heart rate (HR) monitors compared to chest-located HR monitors for training purposes. HR from four wrist-based optical sensor HR products (Fitbit Charge HR, Garmin Vivosmart HR, Apple Watch series 1, Mio Fuse) were compared against a Polar H7 chest strap & RS800cx receiver during nine activities. Two researchers visually observed HR during a protocol incorporating resting, standing, a grocery bag carry, and a 6-stage cycle ergometer protocol that reached maximal HR. Pearson's r and interclass correlations (ICC) in the sample (n=45, mean age=20.22 [SD 2.32]) resulted in the following: Mio Fuse r=.93, ICC=.97; Apple Watch 1 r=.91, ICC=.95; Fitbit Charge HR r=.83, ICC=.91; and Garmin Vivosmart HR r=.74, ICC=.85 (all p's <.001). Bland-Altman plots showed the lowest bias for the Mio (-3.30 bpm), followed by the Apple Watch (-2.82 (SD:14.6) bpm), Garmin (-2.99 (SD:23.9) bpm) with Fitbit having the highest bias (-8.13 (SD:20.6) bpm). No drift in bias was found for any device in successive HR categories (all *p's* >.09). Wrist-based HR monitors were deemed acceptable for fitness classes, though caution should be taken when interpreting any singular visually observed measurement point.

KEYWORDS
Smartwatch; Heart Rate Monitoring; Fitness; Fitness Watch; Validity; Exercise; Cycle Ergometer; Training; Intensity

INTRODUCTION

Over the past ten years participation in group fitness classes has been a method that individuals use to engage in regular physical activity in a social and non-competitive environment. Fitness classes may utilize a range of modalities (i.e., cycling, resistance, step, and Pilates), but cycling has become one of the most practiced activities in fitness centers for people regardless of their physical conditioning level.[1, 2] Participants may monitor their exercise intensity to ensure they are meeting workout goals, avoid over- or under-exertion, and to track changes in fitness over time. Various scales including variations of the Borg Rating of Perceived Exertion (BORG) and OMNI Picture System of Perceived Exertion (OMNI) are used to evaluate intensity levels based on rating physical effort,[3] however objectively assessing the quality of physical activity (PA) can be superior but also challenging. Wrist-based activity monitors that measure heart rate (HR) have been shown to be a practical tool to better understand the quality of PA in directed fitness classes.[4]

The utility of consumer activity trackers has gained mainstream popularity and been highly marketed in medical[5-8] and commercial settings.[3, 9, 10] The wearables market, consisting of smartwatches, chest straps, armbands, head-mounted displays, sensor-embedded clothing, and wrist-secured activity trackers contain various sensor technologies that monitor a variety of biofunctions (i.e., movement, sweat response, pulse oximetry, HR, etc.), but some of these wearables require a second device to see the biofeedback and can be uncomfortable to wear during a high-intensity exercise. In 2016, 102.4 million wearables were shipped with a year-over-year growth of 25.0%.[11] Projected growth of wearables, as reported by the International Data Corporation, was expected to reach 125.5 million shipments in 2017 and grow to 240.1 million by 2021.[12] These projections were an underestimation with double the shipments in 2021 at 533.6 million units.[9] Consumer demand and marketplace competition are driving the number of innovative features, but concerns over which and how biometrics are reported to consumers has prompted the scientific community to assess the validity of features and recommend how much confidence in the metrics should be afforded. The applicability of the reported data, often displayed on a screen, could vary over a variety of activities including light

to vigorous activity and by populations ranging from children to adults. Typically, assessing the validity of either research-grade or commercial-grade physical activity monitors has focused on movement and estimating energy expenditure. Metrics range from accelerometer-derived estimates of light, moderate and vigorous physical activity minutes, steps, GPS tracking, to more recent algorithmic attempts at activity pattern recognition, which remain proprietary in most cases to the manufacturer.[13-15]

HR monitoring is a common feature in wearable devices. Part of the rationale for measuring HR is its strong correlation with oxygen consumption[16] which provides wearables manufacturers an estimation of PA intensity. Prior validation of HR wearable chest strap devices show a strong correlation ($r = 0.99$) when compared to electrocardiogram (ECG).[17, 18] Although well validated, chest strap devices introduce wearability issues such as comfort and are not easily integrated into daily exercise. Therefore, the ability to use a lower-burdening wrist-based HR monitor alternative would be advantageous to implement for guided group exercise classes.

The convergence of several technologies measuring environmental and biometric characteristics into wrist-worn devices has predominantly incorporated optical-based photoplethysmography (PPG) to measure HR by detecting blood volume changes through the skin's surface.[19] The utility of HR measures can be used to infer and quantify the extent of exertion during physical activity, allowing those leading and participating in group exercise to understand their level of effort. As with other biometric measures, HR measures should be independently rigorously evaluated prior to widespread use.[20]

Several consumer-based validity studies have published HR monitoring results.[10, 21-25] These studies have focused on resting[10, 21-24] and exercise modalities including treadmill,[10, 21-25] elliptical,[10] and cycle ergometry[10, 22] using one or multiple stages of increasing intensities. Overall, these studies demonstrate correlations (i.e., $r \geq 0.90$) comparable against electrocardiogram and chest strap devices, but resulted in varying levels of residual error depending on the device or exercise selection ranging from 1.8% to 16.7%.[10, 22-25] Conclusions drawn from these studies reported wrist-based HR measures are more accurate during rest than for exercise[21, 23] and may demonstrate sufficient validity for casual, consumer-based use but the results have a limitation.[10, 21, 23, 25] Although various ranges of HR response have been studied using pre-defined absolute intensities during exercise protocols, upper limits of HR response have not been adequately sampled requiring additional research with younger groups who can attain near maximal HR values.[26]

The purpose of this study was to test the validity of four consumer-based wrist-based HR monitors against a chest-strap monitor using visual inspection across an expanded HR range compared to previous studies to examine validity and bias at higher HRs. Results from this study will inform users on the validity of wrist-based consumer-grade HR monitors for high-intensity exercise and help individuals compare which affordable wrist-based HR monitor is best for them.

MATERIALS AND PROCEDURES

Study Design

The overall study design was cross-sectional using a concurrent validity model that captured visually observed optical HRs from five devices during a nine-stage activity protocol. This included four consumer-available wrist-based optical HR devices with a Polar HR chest strap used as the criterion device. Pretesting was performed to ensure nominal values were obtained against the chest strap over a range of exercise intensities prior to the start of data collection. All procedures were approved by the University's ethical review board prior to completion of any research activities. Written informed consent was collected from each participant prior to the conduct of any research activities (IRB# IRB-2016-05-06-143848).

Participants

A convenience sample ($n = 45$) consisted of undergraduate university students recruited through flyers and classroom announcements. This sample was chosen due to the ability to motivate participants to elicit a near-maximal HR response. Inclusion criteria consisted of college students or university employees between 18 and 30 years of age, being male or female, and belonging to any race or ethnicity. Exclusion criteria included metabolic dysfunction (i.e., type I or II diabetes), implanted pacemaker, and cardiovascular or cerebrovascular conditions. Prior to signing the informed consent, the qualification to be included in the study was determined by passing the Physical Activity Readiness Questionnaire Plus (PAR-Q+) in addition to not taking medications that affect HR or blood pressure.[27] No participant experienced adverse events during the conduct of the study.

Devices

Selection of devices was limited to models available during the first quarter of 2016. The concurrent validity device was a Polar chest strap HR monitor (Model Polar H7 with polar RS800cx watch receiver; Polar Electro Inc., Kempele, Finland). HR measures from this model have been extensively used in other research studies.[17, 28] Studies indicate the Polar chest strap HR monitor exhibits high correlation and low error ($r = 0.99$, ICC > 0.999, error rate = 0.086%) against ECG measurements.[10, 21, 29] The four wrist-based HR devices included the Apple Watch (Apple) (1st generation, OS: 2.82 firmware: 57.11; Apple Inc., Cupertino, CA),

Garmin Vivosmart HR (Garmin) (OS: 2.99, firmware: 93.85; Garmin Corp., Olathe, Kansas USA), Fitbit Charge HR (Fitbit) (OS: 80.77, firmware: 8.13, Fitbit; San Francisco, California USA) and Mio Fuse (Mio) (OS: 49.34, firmware: 1.18; Mio Global; Vancouver, British Columbia Canada). Each device used its own proprietary software. All were based on optical sensor technology to measure HR. For the duration of the study, auto-updates were disabled to maintain device firmware.

Development

The protocol was developed to include equal representation of HR measures along the total range of individuals' responses from resting to near maximum effort using maximum HR ranges (i.e., Astrand formula: max HR = 220 - age).[26] Both free-living stages (i.e., resting, standing, and a simple carry task) and structured exercises (i.e., using a cycle ergometer) were included in the study protocol. Cycle ergometer stages had increasing levels of intensity based upon individual HR ranges **(see Table 1)**. Five participants not included in the analysis were used to collect developmental HR responses to verify measurement procedures, time (i.e., 3 minutes) to establish a steady-state HR measure, device wear locations, and HR response ranges. In our preliminary testing, we found no effect of device location on HR when the manufacturers' instructions were followed. Therefore, device location and fitting were standardized to limit inadvertent pressing of the devices' buttons during the protocol. Since PPG measures HR at the capillary bed level, the authors surmised that counterbalancing devices with different placements was unnecessary compared to other wrist-based variables such as accelerometry, where measures are more susceptible to error due to biomechanical moment arms during movement. Measurement samples were visually taken during a short period (i.e., <10 seconds) with the participant's forearm motionless and parallel to the floor. Visual inspection was used in the methodology to capture HR data to mimic data collected in the field from students' perspectives. Participant burden was 1 hour over 1 visit and no incentive was provided.

Stage	Activity description	Estimated time (min)	Cumulative Time (min)
1	At rest (seated)	1	1
2	Standing desk task reading material	3	4
3	Grocery bag carry with 22 kg in each bag, one in each hand – 10 meter distance, placing bag down at each end.	3	7
4	Cycle (resting) (take measurement when HR<100) no pedaling	3	10
5	Cycle stage 1: target HR (90-120)	3	13
6	Cycle stage 2: target HR (121-135)	3	16
7	Cycle stage 3: target HR (136-150)	3	19
8	Cycle stage 4: target HR (151-175)	3	22
9	Cycle stage 5: target HR (176+) max effort (max 2 min	2	24

Table 1. Activity protocol.
Estimated time includes time to reach steady state, measurement and transfer to next stage. Simultaneous heart rate measures taken during last 10 seconds of each stage.

Protocol

After written informed consent, each participant completed a demographic questionnaire consisting of self-reported age, race, sex and ethnicity with laboratory measures for height (Seca Portable Stadiometer, model 213, Seca Gmbh & co., Hamburg, Germany), weight (weight scale; Seca Gmbh & co., Hamburg, Germany Model: 869 1321004) and wrist circumference (spring-loaded measurement tape; Gulick II, Country Technology, Inc. Gays Mills, WI). Prior to the protocol stages, each participant was fitted with devices in the following order **(See Figure 1)**. On the right arm, the Fitbit was located closer to the hand with the Apple Watch closer to the elbow. On the left arm, the Mio was closer to the hand with the Garmin closer to the elbow. The Polar chest strap was placed around each person's chest and secured per the manufacturer's instructions with the receiver held by a research assistant. Participants remained seated while devices were adjusted until all resting HRs were consistently displayed. During each measurement stage, two research assistants, positioned on either side of the participant, took HRs from all devices simultaneously within ten seconds of ceasing the free-living activity stages (stages 1-3) or while continuously pedaling during the cycle ergometer stages (stages 5-9). One research assistant would take the measurement from one hand, and the other would take the Polar HR monitor and the other two with the Polar HR taken first. If a device's HR measure was missing and or delayed for greater than ten seconds, then the data point was recorded as missing for that device.

Figure 1. Device Placement. Right arm (Fitbit closer to hand, Apple Watch). Left arm (Mio closer to hand, Garmin).

Data Analysis

Demographics were analyzed using descriptive statistics including means, standard deviations, and percentages. Normality was tested using Shapiro-Wilk testing. Missing data was not imputed in the analyses. Device HR output metrics were analyzed using Pearson's correlations with a significance level of 0.05. An a-priori Pearson's correlation value of $r \geq 0.80$ compared to the Polar chest strap was chosen as the acceptability cut-point as used by Gillinov and colleagues.[10] For Interclass-correlations (ICC), a value of above 0.70 was used as a cut-point, and for mean absolute percent error (MAPE), a value lower than 5% was used as reported in other validity studies.[4, 30] Other validity measures included mean absolute error (MAE), calculation of the bias and 95% limits-of-agreement using Bland-Altman plots for comparison to other studies. Secondary analysis of MAPE was stratified by the 9 activities sorted by increasing average HR to assess trends. Change in MAPE across intensity stages were analyzed using multi-variate analysis controlled for sex and race. Data was analyzed using IBM SPSS Statistics, Version 27.0 (Armonk, NY: IBM Corp.).

RESULTS

Forty-five participants were recruited and completed the protocol. Descriptive statistics are provided in **Table 2.** Participants were college-aged (mean: 20.22, Range: 18-28) and were predominantly White (93.0%) and female (82.2%). No participants had tattoos present at the wear location.

Variable	Measure
Age, years, mean (SD)	20.22 (2.32)
Sex	
Female, %	82.2%
Male, %	17.8%
Race	
White n, %	42, 93.3%
Black n, %	2, 4.4%
Asian n, %	1, 2.2%
Wrist size (cm), mean (SD)	15.59 (1.08)
BMI (kg/m^2), mean (SD)	26.07 (18.31)

Table 2. Descriptive Characteristics.
SD: standard deviation; n = 45

Table 3 reports the validity metrics from the study. The mean and standard deviation for HR were similar for all devices in the study. Pearson's correlation coefficients for the devices compared to the Polar HR monitor from strongest to weakest were ordered Mio ($r = 0.93$), Apple Watch ($r = 0.91$), Fitbit ($r = 0.83$), and Garmin ($r = 0.74$), respectively. All devices except the Garmin were higher than the predetermined cut-point of $r = 0.80$. Additional Pearson's correlations between devices showed the strongest association for the Apple and Mio devices ($r = 0.90$), followed closely by the Fitbit and Apple ($r = 0.86$) and Fitbit and Mio ($r = 0.81$), with moderate correlations between the Apple and Garmin ($r = 0.74$), Garmin and Mio ($r = 0.71$) and Fitbit and Garmin ($r = 0.63$). All associations were statistically significant at $p < 0.05$. Interclass correlations for each device were rated higher than the 0.70 cut-point with the same rankings as the Pearson's correlations (all ICC's > 0.85, $p<.001$). MAPE was

relatively high for each device with the Apple Watch and Mio showing mean values at 5% or below. The Garmin showed the highest MAPE (12.07%).

	Criterion: Polar HR	Fitbit	Apple	Garmin	Mio
HR, m(SD)	121.00(33.8)	115.1(36.5)	119.8(34.1)	120.1(30.0)	121(34.2)
Pearson's r	-	.834 (p<.001)	.913 (p<.001)	.739 (p<.001)	.934 (p<.001)
ICC	-	.909 (p<.001)	.954 (p<.001)	.845 (p<.001)	.966 (p<.001)
MAE	-	10.419	5.286	13.627	4.484
MAPE, %	-	8.563	4.386	12.066	4.088
Activity 1	84.2 (14.14)	3.570	3.477	9.396	4.390
Activity 2	92.8 (16.98)	7.738	4.227	12.326	3.272
Activity 3	99.0 (19.51)	8.883	4.415	28.706	6.187
Activity 4	94.3 (15.27)	9.820	4.037	12.144	8.628
Activity 5	112.8 (9.97)	10.185	6.206	10.679	2.165
Activity 6	128.8 (5.26)	10.632	1.456	4.471	2.448
Activity 7	144.9 (6.31)	6.009	4.083	7.060	1.819
Activity 8	162.7 (6.26)	1.040	.569	8.615	.554
Activity 9	178.3 (4.84)	4.076	.632	9.172	.479
p-trend	-	p=.40	p=.09	p=.42	p=.036
Bias, m(SD)	-	-8.134(20.60)	-2.816(14.57)	-2.987(23.94)	-3.303(25.31)
LOA 95% CI	-	(32.25, -48.52)	(25.74, -31.37)	(43.94, -49.91)	(24.12, -25.22)

Table 3. Validity and agreement measures between devices.
m(SD) = mean (standard deviation), ICC = interclass-correlation, MAE = mean absolute error
MAPE = mean absolute percent error. LOA = limits of agreement.

Examining MAPE by activity is visually depicted in **Figure 2**. In general, we see lower MAPE at rest (Activity 1) and at higher activity levels (Activity 8 and 9). Higher MAPE was seen in the Garmin, especially during Activity 4, followed by the Fitbit, which increased in moderate to higher HR activities then reduced at higher intensities. Trends for the Fitbit, Apple, and Garmin were flat as intensity increased but the Mio had a trend toward a smaller MAPE as intensity increased (p = .036). To provide a real-world metric to describe the error, time-aligned HRs from each device were compared to the Polar criterion to establish a percentage where readings were within 5 bpm of each other. Percentages were highest (i.e., higher is better) for the Mio at 85.2% and Apple Watch at 82.2%, followed by the Fibit at 66.6% and the lowest for the Garmin at 57.4%. As an example, one would interpret the data as: 85.2% of the time, the values from the Mio device were within 5 bpm of the Polar device.

Figure 3 reports Bland-Altman plots with 95% limits of agreement (LOA) for each device compared to the Polar chest monitor. Plots illustrate the distribution of HR residuals across the observed HR range for all 45 participants with HR estimates (y-axis: wrist-based device subtracted from the Polar (bpm)) relative to the mean of the two comparison devices (x-axis: mean of Polar and respective devices' (bpm)). Results show the LOA have the tightest clustering for the Mio (LOA: 24.12, -25.22), followed by the Apple Watch (LOA: 25.74, -31.37), Fitbit (LOA: 32.25, -48.52), and Garmin (LOA: 43.94, -49.91). Using all data, each device was negativly biased, underestimating HR compared to the Polar, with the lowest bias for the Mio (-.55), followed by the Apple Watch (-2.82), Garmin (-2.99) with the Fitbit having the highest bias (-8.13). Exploratory analysis was performed using a subjective author-determined value of +/- 20 bpm difference to mimic ignoring a very low or high HR value reported to the user. This resulted in the lowest adjusted bias values for the Mio (-.16), followed by the Garmin (-.28), Apple Watch (-.43) with the Fitbit having the highest bias (-1.71). Each device had the majority of their residuals around 0, though error values were more noticable for the Fitbit and the Garmin devices. In general, the shape of the residuals for all devices generally showed more error near the center of the HR distribution compared to the tails, with the majority of error approximately in the 90-130 bpm range. Multivariate regression showed all devices had slopes that were not different than 0 (all p's > 0.09), meaning no systematic bias was found as HR response increased. Further modeling controlling for race and/or sex showed no systematic bias (all p's > .10).

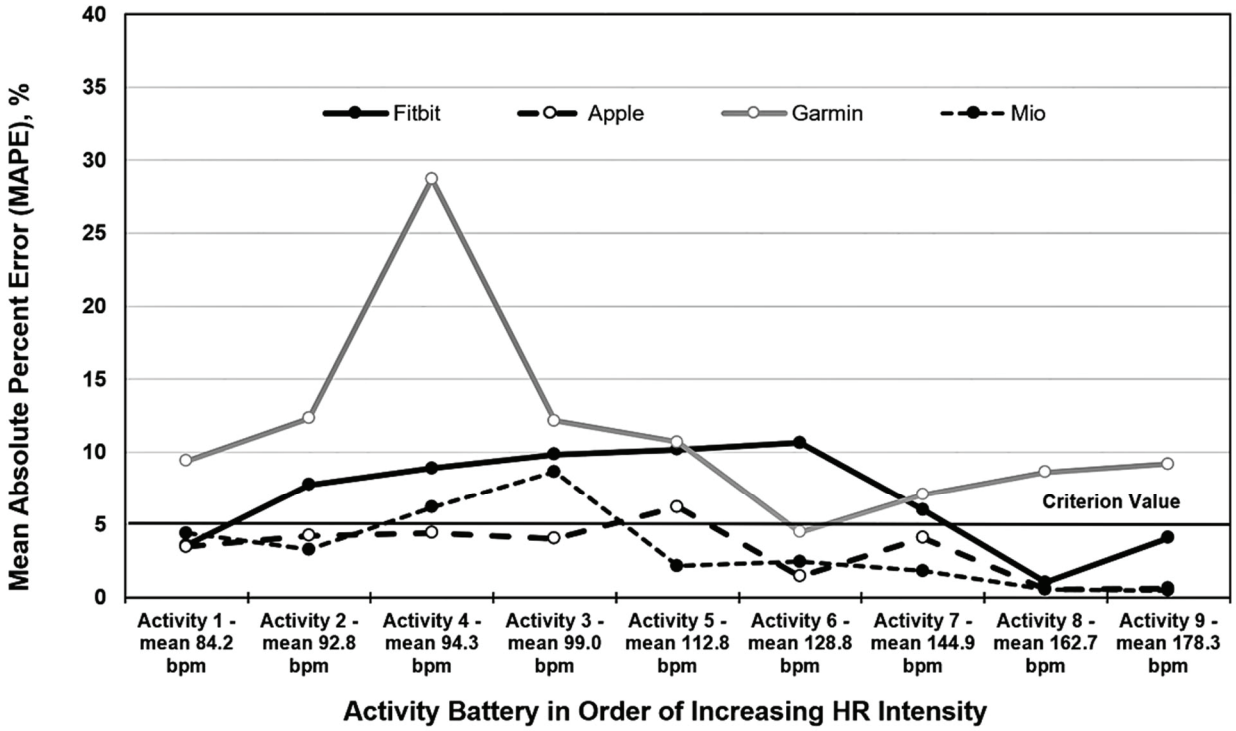


Figure 2. Mean absolute percent error for heart rate stratified by activity stage and device (Fitbit Charge HR, Garmin Vivosmart HR, Apple Watch series 1, and Mio Fuse compared to Polar H7 chest strap and RS800 cx receiver criterion device. Cycle values represent targeted HR ranges during cycle ergometry stages.

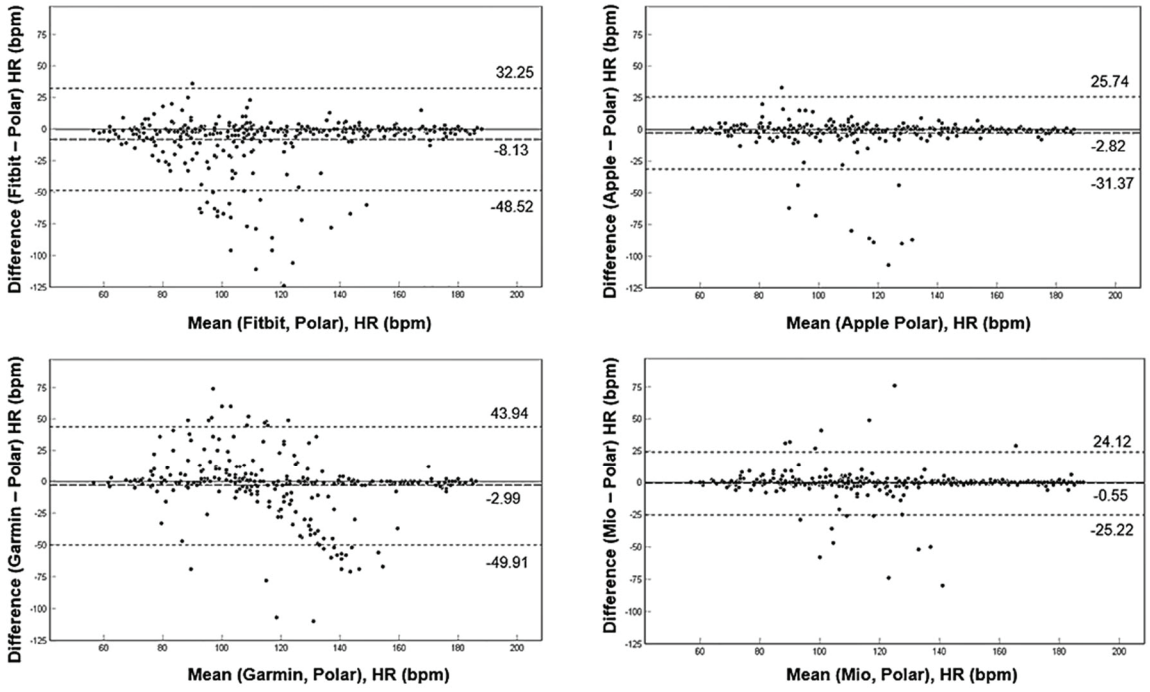


Figure 3. Bland Altman plots of each consumer wrist-based heart rate monitors (Fitbit Charge HR, Garmin Vivosmart HR, Apple Watch series 1, Mio Fuse) compared to Polar H7 chest strap and RS800 cx receiver showing overall bias (solid line) and 95% limits of agreement.

DISCUSSION

This study examined the concurrent validity of four commercially available, consumer-based, wrist-worn, HR activity monitors to inform group exercise leaders, coaches, personal trainers, and participants on the validity of HR measures across a wide range of HRs. Overall, our findings suggest moderately acceptable validity for the wrist-based monitors compared to the chest-based criterion with certain models exhibiting higher error than others. The four monitors showed moderate to high correlations and ICC's but MAPE varied with HR intensity. The device with the lowest overall error, bias, and correlations was the Mio Fuse closely followed by the Apple Watch with higher error for the Fitbit and Garmin devices. Overall, the Mio and Apple Watch had comparable and lowest bias, highest correlation, and narrowest limits of agreement.

This article specifically targets the validity of wrist-based devices at near maximum HR. All devices were found to have no slope regarding bias as HR intensity increased instilling confidence that higher intensities will not be over or underestimated more than lower HR ranges. However, random variability of HR observations outside of 5 bpm was endemic for all devices. This was most prevalent in the Garmin device but relevant for all devices, especially during moderate-intensity exercise. We hypothesize several sources of error. One is the onset of sweat or change in perfusion of tissues may introduce artifacts through altering the optical sensor reflection or changes in skin permeability.[31] Error may also be influenced by nitric oxide release, causing vasodilation that alter skin thickness.[31]

Our results corroborated other authors' findings with similar device iterations of their time. In Wang et al., 2017, several wrist-worn devices were compared to ECG during a treadmill protocol. They found the Apple Watch series 1 correlation to be r = 0.91, which was identical to our findings (r = 0.91) with comparable results for the Mio Fuse (r = 0.91 compared to the present study findings: r = 0.93). They also reported a wide 95% LOA for all measured devices including the Fitbit Charge HR (range = 73.0), while we found a higher range (range = 93.9).[21] An article by Cadmus-Bertram et al., 2017 reported a similar unacceptable 95% LOA for the Fitbit HR (bias: -2.5, range: 77.0) and Mio Fuse (bias: 1.8, range: 48.5) during exercise, though improved at rest.[23] Similar to Cadmus-Bertram et al., our findings also showed a large negative bias in the Fitbit device (bias = -8.13 (SD 20.6)) with a wide 95% LOA. Stahl et al., 2016 reported acceptable findings during walking and running treadmill activities with all correlations > 0.92.[24] These values were higher compared to the other studies' correlations, including ours.[10, 21] In Stahl et al., the authors analyzed several devices including the Mio Alpha (MAPE = 4.6%, r = 0.93) and the Fitbit Charge HR against a Polar RS400 chest strap (MAPE = 6.2%, r = 0.93). Their study found lower error rates compared to the current study which may be attributed to activity selection and intensities observed.[24] Gillinov et al., 2017 used a criterion of r > 0.80 as a measure of acceptability using running, biking, and elliptical stages that were continuously measured and downloaded for comparison.[10] Across activities, only the Apple Watch, TomTom Spark and Garmin Forerunner were acceptable where the Fitbit Blaze and Scosche Rhythm+ were not.[10] Using their criteria, only the Garmin (r = 0.74) would not reach acceptability in our findings. Their results showed the Apple Watch had the lowest MAPE (4.1%) which was similar to our findings (MAPE = 4.39%). Shcherbina et al., 2017, also reported the lowest amount of error for the Apple Watch compared to ECG when comparing six consumer devices including a Fitbit Surge and Mio Alpha.[22] Overall, other studies reported that HR measured on consumer devices were more accurate at rest, had reasonable error ratings, and generally wide 95% LOA when using them for exercise.[10, 21-24] Potential differences in studies using identical devices may be attributed to the type of measurement method used (i.e., downloaded vs. visual observed), inter-device reliability, differences in activity protocol, or different versions of firmware/software that was used. Studies should report the firmware and operating system version of devices so comparisons can be made in future studies.

Overall, the results of this study are consistent with other studies indicating that wrist-worn consumer-level optical HR monitors may provide an adequate measure of HR for rest and exercise during recreational uses but warrant caution in clinical applications for those with cardiac conditions.[10, 21, 22, 32, 33] In comparison to more recent studies,[25, 32, 33] the validity has not improved over older models; thus variability may be more inherent in the measurement technology or wear and that major improvements may not be realized until newer technologies are developed. This implies that newer, more expensive devices are not superior to older iterations of the same product. They offer the same level of HR monitoring accuracy as the newest devices but at a more affordable price. At the time of this study, there was little difference between the prices of the devices used making each of them a good choice for affordable HR monitoring during exercise, or at rest.

There were several limitations in the current study. This study was designed as a concurrent validity study and did not use the gold standard criterion, ECG, to assess HR measures. However, Polar devices have been shown to be highly correlated with ECG measures (r = 0.99).[28] We did not test for HR max, instead used the participants' calculated HR max. We felt this was acceptable with the younger population, however, if an older population or special population were to be tested, calculated HR max may not be warranted. Another limitation was that only one sample from each device model was used in this study. Therefore, we could not perform reliability analyses. The model year of our devices was from 2016 and do not reflect the potential improvements of current software or hardware changes that are more prevalent at the time of this publication. PPG sensor technology has not drastically changed and we surmise that our findings should be relevant for current devices using this technology. Visual

inspection was the primary measurement method used in this study to mimic end-user experience. Although this method introduces error, it has been used in several other HR validity studies.[10, 21, 23, 24] It is important to note that the results from this study can only be generalized to HR measures taken with the wrist in a steady position such as when one visually inspects their device and not with the wrist moving which likely introduces other HR artifacts.[34] Despite these limitations, the strengths of this study include a sample size that was comparable to other widely cited validity studies and exercise protocols with a sample that allowed for assessment of a wider range of HR responses.

CONCLUSION

This study replicates but also extends findings from other authors on the use of wrist-based HR monitors to near-maximum HR values for PA. With newer models, we cannot expect a large improvement in measurement accuracy, but a shift may be realized as sensor technologies advance. Fitness professionals can utilize wrist-located devices and still maintain a level of accuracy appropriate for recreational use. A trade-off in accuracy and error compared to ease of use may be relevant for personal trainers and fitness coaches to incorporate an efficient measure of intensity during exercise classes.

ACKNOWLEDGEMENTS

We thank the following student research assistants for data collection and development activities: Natalie Sealover, Allie M. Hoover, Heather M. Puleo, Carson McGinn, Quinten Miller, Katherine Kirven, Anna Helm, Cameron Helmer, Katherine White, Tyler Bailey, and Samuel Greene.

REFERENCES

1. Wickham JB, Mullen NJ, Whyte DG, Cannon J. Comparison of energy expenditure and heart rate responses between three commercial group fitness classes. *Journal of Science and Medicine in Sport*. Jul 2017;20(7):667-671. *doi:10.1016/j.jsams.2016.11.012*
2. Chavarrias M, Carlos-Vivas J, Collado-Mateo D, Perez-Gomez J. Health Benefits of Indoor Cycling: A Systematic Review. *Medicina (Kaunas)*. Aug 8 2019;55(8) *doi:10.3390/medicina55080452*
3. Canario-Lemos R, Vilaca-Alves J, Moreira T, et al. Are Heart Rate and Rating of Perceived Exertion Effective to Control Indoor Cycling Intensity? *International Journal of Environmental Research and Public Health*. Jul 4 2020;17(13)*doi:10.3390/ijerph17134824*
4. Brazendale K, Decker L, Hunt ET, et al. Validity and Wearability of Consumer-based Fitness Trackers in Free-living Children. *International Journal of Exercise Science*. 2019;12(5):471-482.
5. Mertz L. Convergence revolution comes to wearables: Multiple advances are taking biosensor networks to the next level in health care. *IEEE Pulse*. Jan-Feb 2016;7(1):13-7. *doi:10.1109/MPUL.2015.2498475*
6. Mombers C, Legako K, Gilchrist A. Identifying medical wearables and sensor technologies that deliver data on clinical endpoints. *British Journal of Clinical Pharmacology*. Feb 2016;81(2):196-8. *doi:10.1111/bcp.12818*
7. Munos B, Baker PC, Bot BM, et al. Mobile health: The power of wearables, sensors, and apps to transform clinical trials. *Annals of the New York Academy of Sciences*. Jul 2016;1375(1):3-18. *doi:10.1111/nyas.13117*
8. Pellanda EC, Pellanda LC. Primordial prevention and wearable health devices: The wearables in cardiology. *Arquivos Brasileiros de Cardiologia*. Jun 2016;106(6):455-6. *doi:10.5935/abc.20160094*
9. Llamas R, Shirer M, Ubrani J. Wearables Deliver Double-Digit Growth for Both Q4 and the Full Year 2021, According to IDC. International Dara Corporation (IDC). June, 29th, 2022, Updated March 9th, 2022. *https://www.idc.com/getdoc.jsp?containerId=prUS48935722*
10. Gillinov S, Etiwy M, Wang R, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Medicine and Science in Sports and Exercise*. Aug 2017;49(8):1697-1703. *doi:10.1249/MSS.0000000000001284*
11. IDC. *Wearables aren't dead, they're just shifting focus as the market grows 16.9% in the fourth quarter, according to IDC*. 2017. *IDC Worldwide Quarterly Wearable Device Tracker*. March 2nd, 2017. Accessed Accessed: September 4th, 2017. *https://www.idc.com/getdoc.jsp?containerId=prUS42342317*
12. Llamas R, Ubrani J, Shirer M. *Worldwide wearables market to nearly double by 2021, according to IDC*. 2017. *Worldwide Quarterly Wearable Device Tracker*. June 21, 2017. Accessed Accessed: September 28th, 2017. *https://www.idc.com/getdoc.jsp?containerId=prUS42818517*
13. Tudor-Locke C, Sisson SB, Lee SM, Craig CL, Plotnikoff RC, Bauman A. Evaluation of quality of commercial pedometers. *Canadian Journal of Public Health*. 2006;97 (Suppl. 1):S10-S16. NOT IN FILE.
14. Welk GJ, McClain JJ, Eisenmann JC, Wickel EE. Field validation of the MTI Actigraph and BodyMedia armband monitor using the IDEEA monitor. *Obesity (Silver Spring)*. Apr 2007;15(4):918-28. *doi:10.1038/oby.2007.624*
15. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *Journal of Medical Engineering and Technology*. Apr 2017;41(3):208-215. *doi:10.1080/03091902.2016.1253795*

16. Bernard T, Gavarry O, Bermon S, Giacomoni M, Marconnet P, Falgairette G. Relationships between oxygen consumption and heart rate in transitory and steady states of exercise and during recovery: Influence of type of exercise. *European Journal of Applied Physiology and Occupational Physiology*. 1997;75(2):170-6.
17. Laukkanen RM, Virtanen PK. Heart rate monitors: state of the art. *Journal of Sports Sciences*. Jan 1998;(Suppl. 16):S3-7. *doi:10.1080/026404198366920*
18. Gregoski MJ, Mueller M, Vertegel A, et al. Development and validation of a smartphone heart rate acquisition application for health promotion and wellness telehealth applications. *International Journal of Telemedicine and Applications*. 2012;2012:696324. *doi:10.1155/2012/696324*
19. Silva I, Lee J, Mark R. Photoplethysmograph quality estimation through multichannel filtering. *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*. 2011;2011:4361-4. *doi:10.1109/IEMBS.2011.6091082*
20. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral and Physical Activity*. Dec 18 2015;12:159. *doi:10.1186/s12966-015-0314-1*
21. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. *Journal of the American Medical Association Cardiology*. Jan 01 2017;2(1):104-106. *doi:10.1001/jamacardio.2016.3340*
22. Shcherbina A, Mattsson CM, Waggott D, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*. May 24 2017;7(2):E3. *doi:10.3390/jpm7020003*
23. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The accuracy of heart rate monitoring by some wrist-worn activity trackers. *Annuals of Internal Medicine*. Apr 18 2017;166(8):610-612. *doi:10.7326/L16-0353*
24. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *British Medical Journal Open Sport and Exercise Medicine*. 2016;2(1):e000106. *doi:10.1136/bmjsem-2015-000106*
25. Jagim AR, Koch-Gallup N, Camic CL, et al. The accuracy of fitness watches for the measurement of heart rate and energy expenditure during moderate intensity exercise. *The Journal of Sports Medicine and Physical Fitness*. Feb 2021;61(2):205-211. *doi:10.23736/S0022-4707.20.11151-4*
26. Oja P, Partanen T, Teraslinna P. The validity of three indirect methods of measuring oxygen uptake and physical fitness. *The Journal of Sports Medicine and Physical Fitness*. Jun 1970;10(2):67-71.
27. Warburton DER, Jamnik VK, Bredin SS, Gledhill N. The Physical Activity Readiness Questionnaire for everyone (PAR-Q+) and Electronic Physical Activity Readiness Medical Examination (ePARmed-X+). *Health and Fitness Journal of Canada*. 2011;4(2):3-23.
28. Giles D, Draper N, Neil W. Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. *European Journal of Applied Physiology*. Mar 2016;116(3):563-71. *doi:10.1007/s00421-015-3303-9*
29. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart rate variability recording with smart phone photoplethysmographic, Polar H7 chest strap and electrocardiogram methods. *International Journal of Sports Physiology and Performance*. Mar 14 2017:1-17. *doi:10.1123/ijspp.2016-0668*
30. Navalta JW, Ramirez GG, Maxwell C, Radzak KN, McGinnis GR. Validity and Reliability of Three Commercially Available Smart Sports Bras during Treadmill Walking and Running. *Scientific Reports - Nature*. Apr 30 2020;10(1):7397. *doi:10.1038/s41598-020-64185-z*
31. Jeong IC, Yoon H, Kang H, Yeom H. Effects of skin surface temperature on photoplethysmograph. *Journal of Healthcare Engineering*. 2014;5(4):429-38. *doi:10.1260/2040-2295.5.4.429*
32. Navalta JW, Montes J, Bodell NG, Salatto RW, Manning JW, DeBeliso M. Concurrent heart rate validity of wearable technology devices during trail running. *PLoS One*. 2020;15(8):e0238569. *doi:10.1371/journal.pone.0238569*
33. Fuller D, Colwell E, Low J, et al. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR Mhealth Uhealth*. Sep 8 2020;8(9):e18694. *doi:10.2196/18694*
34. Zhu S, Tan K, Zhang X, Liu Z, Liu B. MICROST: A mixed approach for heart rate monitoring during intensive physical exercise using wrist-type PPG Signals. *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*. 2015;2015:2347-50. *doi:10.1109/EMBC.2015.7318864*

ABOUT STUDENT AUTHORS

Korey Little is an Exercise Science Major in the Department of Health and Human Performance at the College of Charleston and is expected to graduate with his B.S. in Exercise Science in May 2023. He plans to work in medical research prior to obtaining a DPT.

PRESS SUMMARY

Heart rate is one of the most important things to monitor during exercise as it provides a better understanding of what is happening inside of the body and how hard a person is working. In group exercise classes, the way coaches and trainers control the intensity is usually a combination of using a preplanned exercise program and music. Adding the ability to measure heart rate

would greatly increase the quality of exercise for everyone involved. Typical chest-located heart monitors can be intrusive and difficult to fit comfortably. Having the ability to wear a heart rate monitor on the wrist affords reasonable accuracy and provides more access to group exercise classes or personal training sessions. We tested four different wrist-based heart rate monitors (Mio fuse, Apple watch 1, Fitbit Charge, Garmin Vivosmart) on 45 college-aged participants and found that the Mio Fuse was the most accurate and closest to a medical-grade heart rate monitor followed closely by the Apple watch 1. Most importantly, there was no bias in heart rate as intensity increased to maximum values in all devices. All the watches were accurate enough to be used, but we found that the initial reading may be incorrect and a second or third look may be necessary to increase confidence in observed heart rates.

Retributive Attitudes and Perceptions of Police Use of Excessive Force

Amelia Collins, Sherah L. Basham, & Rick Dierenfeldt*
Department of Social, Cultural, and Justice Studies, University of Tennessee at Chattanooga, Chattanooga, TN

https://doi.org/10.33697/ajur.2023.089

Student: amelialynncollins@gmail.com
Mentors: sherah-basham@utc.edu; ricki-dierenfeldt@utc.edu*

ABSTRACT
Public opinions of police use of force vary widely. Previous studies, however, have framed their examinations around the factors that influence support of police use of force in general, as compared to a focus on excessive force. This study utilized linear regression to examine the relationship between perceptions of police use of excessive force and retributive attitudes. The study employed a sample of 5,527 respondents from the American National Election Studies (ANES) 2020 Time Series Survey. Findings indicated that respondents' perceptions of the frequency of police use of excessive force depend on their retributive attitudes. The more retributive one's attitude, the less often they perceived the police to use too much force. Similarly, the more conservative one's political ideology, the less frequently they perceived police used excessive force. Perceptions of police excessive force also vary across demographics.

KEYWORDS
Retributiveness; Death Penalty; Police Use of Force; Police Excessive Force

INTRODUCTION

Police use of force has garnered significant attention in recent years due to high-profile incidents of excessive force. Although research demonstrates that police seldom use force[1, 2] and excessive force is even more rare, such instances spark public attention.[3] Literature about public perception and attitudes toward police use of force has grown.[2,4] Yet, much of this work focuses on support for police use of force in general terms (e.g., lawful use of force).[5,6] However, as argued by Drakulich et al., instances of police use of excessive force draw public concern, evincing a need to examine perceptions of the frequency with which police use force beyond what is permitted by law.[7]

Retributive theory is a framework through which criminal justice actions can be viewed as responses to human behavior. While much has been written on support for police use of force, less is known about the factors that shape public perceptions of how often police use excessive force. Are there retributive attitudes that influence perspectives of police excessive force? Do one's political ideologies and demographics impact their perceptions of how frequently police use too much force? Utilizing the 2020 American National Election Studies (ANES) Study, this study explores the relationship between retributive attitudes and perceptions of police excessive force. We begin with an overview of police use of force and retributive theory and follow with a review of factors that influence perceptions of police excessive force.

LITERATURE REVIEW

Within the field of policing, the use of force is the amount of force applied by a police officer to secure compliance through physical force or verbal commands.[8] As noted by several researchers, the application of physical force by the police is relatively rare.[1, 2] Acceptable uses of force derive from procedural justice, the fairness of the process through which police make decisions and exercise authority.[9,10] Legal statutes, professional standards, and societal expectations guide appropriate conduct. However, recent high-profile abuses illustrate that some police officers use levels of force that violate these social and legal standards. According to Gerber and Jackson, police use of force becomes excessive when the amount of force surpasses what is necessary to control the situation and address the seriousness of the threat.[4] Despite its rarity, excessive force is one of the most common forms of police misconduct and, therefore, is intensely regulated.[5, 11]

Retribution

Retributive theory is a theory of punishment that focuses on the concept of just deserts. It is one of the significant philosophical theories that seek to justify punishment as a response to criminal behavior. According to retributive theory, individuals who commit crimes deserve to be punished proportionally to the harm they have caused, regardless of the potential consequences of the punishment or its future deterrent effect.[12, 13] Retributive theory is composed of several tenets, including deserved

punishment, just deserts, and moral equality. Retributivists argue that punishment is morally justified because it is what wrongdoers deserve as a consequence of their actions. The punishment should be proportional to the severity of the crime committed. In other words, the punishment should fit the crime. Just deserts refers to the idea that individuals are morally responsible for their actions and should be held accountable. Punishment is viewed and treated as mechanism through which to restore moral balance and uphold the principles of justice. The retributive theory also argues for moral equality in punishment, which implies that those who commit the same crime should receive the same punishment. This concept aims to treat individuals as equals before the law and avoid arbitrary or discriminatory practices.

Critics of retributive theory point out that it can lead to harsh and inflexible punishment without considering the potential for rehabilitation or addressing the root causes of criminal behavior. It is also challenging to determine a universally agreed-upon level of proportionality for each crime, leading to discrepancies in sentencing.[12] In modern legal systems, retributive theory often coexists with other theories of punishment, such as deterrence, rehabilitation, and restorative justice. The balance between these theories varies depending on the jurisdiction and the prevailing societal values.[13]

Retributiveness and Excessive Force

Support for the death penalty is frequently used in research to measure retributiveness or punitiveness.[6,15,16,17] As such, it is also used to indicate social and criminal justice perspectives.[18,19] Exum approached the relationship between police use of force and the death penalty from a legal viewpoint. When police use excessive force, rights protected under the Fourth Amendment are violated. Exum argued that by viewing police use of lethal force as a form of punishment and "the administration of the death penalty on the streets" (p.988), then the Fourth Amendment is invoked. According to the Fourth Amendment, people are protected from unreasonable search and seizure. Thus, the Fourth Amendment is a lens through which to scrutinize the (un)reasonableness of lethal force by police.[20]

However, as explained by Mourtgos and Adams, "evaluations of police use-of-force are often examined from the viewpoint of a prevailing community standard, not that of a purely legal viewpoint (p.872)."[5] Therefore, what the public views as excessive may not align with prevailing legal standards. According to Cullen et al., embracing a more retributive or punitive crime ideology, which involves advocating an "eye-for-an-eye" approach to crime control, will likely influence perceptions of police-citizen encounters.[21] Research has indicated that those who favor the death penalty tend to be more approving of the use of force by law enforcement.[6,22] Applying the retribution theory, Silver and Pickett argue that this could be the result of embracing perspectives of just deserts.[6] Silver and Pickett studied the connection between police use of force and support for the death penalty through a politicized policing outlook, concluding that there are predictors between attitudes toward police use of force, excessive force, and the likelihood of supporting the death penalty as a measure of retributiveness.[6]

Public perception of police use of force is premised upon various individual factors. Indicators of public perception of police use of force have been found in race/ethnicity and political affiliation. Studies consistently find that minority racial and ethnic groups are less likely to support police use of force than whites.[5,6,19] Specifically related to perceptions of excessive force, Black and Latinx Americans perceive that the police use excessive force at significantly higher rates.[7] They also found that racial prejudice was a significant predictor of support for police use of excessive force across political groups.[5] Scholars have also accounted for various demographic factors, finding that those with more income and higher education were less likely to perceive excessive force as a common occurrence.[7]

Based on the review of the extant literature, we hypothesize the following:

H_1: Perceptions of the frequency of police use of excessive force will be inversely related to retributive attitudes.
H_2: Liberal political ideologies will correspond with perceptions that excessive force is used more frequently by police.

METHODS AND PROCEDURES

This study utilized secondary data from the American National Election Studies (ANES) 2020 Times Series Study.[23] Data were gathered through various methods, including phone, web, and video surveys and interviews. Pre-election surveys and interviews were administered between August 18, 2020, and November 3, 2020. Post-election data were collected between November 8, 2020, and January 4, 2021. The final ANES 2020 study sample comprised 8,280 pre-election respondents and 7,449 post-election responses.[23] For the purposes of this study, only those who participated in the pre-and post-election ANES surveys were eligible for inclusion (n=7,449). Any cases without data for the variables of interest were removed from the sample. The final sample was composed of 5,527 respondents.[A]

[A] See American National Election Studies (2021) for a complete explanation of ANES methods and sampling.

Variables

The dependent variable was the perceived frequency with which police use excessive force. In the post-election survey, respondents were asked, "How often do you think police officers use more force than is necessary?" A five-item scale was provided with higher values indicating a high frequency of police excessive force. About 54.5% of respondents reported perceiving that the police used too much force half the time or more. The average response was 2.82, illustrating that the average survey respondent believed that more than half the time, the police use more force than necessary.

The study's primary independent variable was retributiveness. Previous scholars have demonstrated that retributive attitudes are related to endorsing the death penalty.[6, 17] Therefore, using the pre-election survey, respondents were asked, "Do you favor or oppose the death penalty for persons convicted of murder?" Responses ranged from Oppose strongly (less retributive), to favor strongly (more retributive) on a four-item scale, with 62.1% reporting that they favored the death penalty to some degree. The average response was 2.85, illustrating that the average survey respondent supported the death penalty at some level.

The second independent was political ideology. In the pre-election survey, respondents were asked where they would place themselves on a seven-point liberal-conservative scale where higher values demonstrate conservative political ideology. Approximately 39.1% reported being conservative to some degree. The study controlled for a variety of respondent demographics, including age, self-reported sex, sexual orientation, race/ethnicity, total family income, education, and political ideology. Using generational categories employed by Pew Research, age was measured in 5 generational categories.[24] Generation Z represents respondents born between 1997-2002, ages 18 to 23. Millennial represents respondents born between 1981-1996, ages 24-39; Generation X represents respondents born between 1965-1980, ages 40-55; Baby Boomers represent respondents born between 1946-1964, ages 56-74, and the Silent Generation represents respondents born 1928-1945, ages 75 and higher. As the largest category (29%, n=1,601), Baby Boomers were selected as the reference category.

The ANES survey asked, "What is your sex?" Male and female response categories were provided. Male was coded "0," and Female was coded "1." The majority of the respondents were male (50.6%, n=2,795). Respondents were also asked about their sexual orientation: "Do you consider yourself to be heterosexual or straight, homosexual or gay [or lesbian], or bisexual?" Four response categories were listed: Heterosexual or Straight, Homosexual or Gay (or Lesbian), Bisexual, or Something Else. Due to few responses in some response groups, the categories were condensed into two categories: Heterosexual or Straight (0) and Not Heterosexual (1). The majority of the respondents reported that they were heterosexual/straight (93.1%, n=5,145). Race/ethnicity was measured as White/Non-Hispanic, Black/Non-Hispanic, Hispanic, Asian or Native Hawaiian/other Pacific Islander/Non-Hispanic alone, Native American/Alaska Native or Other race/Non-Hispanic alone, and Multiple Races/Non-Hispanic. Due to low responses in some categories, they were collapsed into four categories: White/Non-Hispanic (70%, n=3,869), Black/Non-Hispanic (9.0%, n=496), Hispanic (11.2%, n=619), or Other (9.8%, n=543). White/Non-Hispanic was the referent category.

Finally, income was a summary variable from the pre- and post-survey, representing total family income as a 22-item scale. A higher rating demonstrates a higher income. In the pre-election survey, respondents were asked about their highest income level on an eight-item scale; higher values represent a higher level of education. About 39.9% reported having a bachelor's degree or higher. See **Table 1** for all descriptive statistics.

RESULTS

Ordinary least squares regression was used to examine the research question, *What is the relationship between relationship between perceptions of police use of excessive force and retributive attitudes?* The dependent variable, perceived frequency of police excessive force, was regressed on 13 items in total. The results are presented in **Table 2**. The model was fit to the ANES data using survey weights. The model was significant, R^2=0.202, F(13)= 184.747, p<.001. Variance inflation values ranged from 1.01 to 1.49; thus, multicollinearity was not an issue. The independent variable, retributiveness, was significant, demonstrating a significant relationship between respondent perceptions of the frequency of police use of force and retributiveness. As support for the death increases, perceptions of how often the police use excessive force decrease (b=-0.066, SE=0.011, p<0.001). This outcome supports our first hypothesis, that perceptions of the frequency of police use of excessive force would be inversely related to retributive attitudes.

Considering the second independent variable, political ideology (b=-0.187, SE=0.008, p<0.001) was a significant predictor of the perception of police excessive force. Respondents who were more liberal perceived that police use too much force more frequently. Based on the standardized beta coefficients, political ideology (β=0.320) has the strongest effect on perceptions of police use of excessive force, over four times stronger than that of retributive attitudes (β=0.078). This finding supports our

second hypothesis, that liberal political ideologies will correspond with perceptions that excessive force is used more frequently by police.

Variable	Mean	SD	Range	N	%
Belief that the police use too much force	2.82	0.959	1 - 5	5527	
Retributiveness: Support death penalty	2.85	1.134	1 - 4	5527	
Political Ideology	4.13	1.636	1 - 7	5527	
Age by 5 Generations					
Generation Z: 1997 through 2002				488	8.80
Millennial: 1981 through 1996				1596	28.90
Generation X: 1965 through 1980				1505	27.20
Baby Boomers: 1946 through 1964				1601	29.00
Silent: 1928 through 1945				338	6.10
Sex					
Male				2795	50.60
Female				2732	49.40
Sexual Orientation					
Heterosexual				5145	93.10
Non-Heterosexual					
Homosexual, Gay, or Lesbian				158	2.90
Bisexual				166	3.00
Something Else				58	1.00
Race/Ethnicity					
White/non-Hispanic				3869	70.00
Black/non-Hispanic				581	8.80
Hispanic				496	9.00
Other					
Asian, Native Hawaiian, Pacific Islander				225	4.10
Native American, Alaska Native, Other				100	1.80
Multiple Races				218	3.90
Income	13.92	6.23	1 -22	5527	
Education	4.17	2.075	1 - 8	5527	

Table 1. Univariate Statistics N=5,527

The demographic control variables provided a variety of outcomes. In terms of age, two of the four generational categories presented significant relationships. In comparison to the Baby Boomer Generation (age 56-74), respondents in both Generation Z (age 18-23) (b=0.252, SE=0.043, p<0.001) and the Millennial Generation (b=0.194, SE=0.029, p<0.001) have perceptions that the police use excessive force more frequently than Baby Boomers. However, neither Generation X (b=0.035, SE=0.029) nor the Silent Generation (b=-0.0795, SE=0.048) respondents' perception of police use of excessive force differed significantly from the Baby Boomer Generation.

When considering sex, female respondents perceive that the police use excessive force more frequently (b=0.70, SE=0.022, p<0.001) than males when holding all else constant. Consistent with prior research, female survey respondents have a 16.8% reduction in the odds of favoring the death penalty when holding all else constant. As for sexual orientation, respondents who identify as non-heterosexual perceive that the police use excessive force more frequently than heterosexuals (b=0.302, SE=0.044, p<0.001). Black (b=0.723, SE=0.039, p<0.001), Hispanic (b=0.405, SE=0.035, p<0.001), and Other race (b=0.353, SE=0.037, p<0.001) respondents perceived that police use too much force significantly more often than White respondents. Finally, focusing on income and education, as respondents' income (b=-0.016, SE=0.002, p<0.001) and education (b=-0.037, SE=0.006, p<0.001) increased, their perception of the frequency of police use of excessive force decreased.

Variable	*b*	SE
Retributiveness	-0.066***	0.011
Political Ideology	-0.187***	0.008
Age: Generation Z	0.252***	0.043
Age: Millennial	0.194***	0.029
Age: Generation X	0.035	0.029
Age: Silent	-0.079	0.048
Sex: Female	0.079***	0.022
Sexual Orientation: Non-Heterosexual	0.302***	0.044
Race: Black, non-Hispanic	0.723***	0.039
Race: Hispanic	0.405***	0.035
Race: Other	0.353***	0.037
Income	-0.016***	0.002
Education	-0.037***	0.006
Constant	3.877***	0.058

Note: *p <.05. **p <.01. ***p <.001.

Table 2. Linear Regression (N=5527)

DISCUSSION

This study examined the relationship between perceptions of police use of excessive force and retributiveness. Prior studies researching the connection between police use of force focus on an array of factors such as warranted police use of force, legal concerns about the use of force, and general support for police use of force. These studies leave room for questions concerning further connections in the relationship between public perceptions of how often the police use too much force and the retributive nature of such actions.

The 2020 ANES captures public perceptions of issues such as police excessive force and retributive attitudes. This study explored retribution by using support for the death penalty as a predictor of perceptions of how often police use excessive force, resulting in several important findings. First, we found that retributiveness predicts perceptions of the frequency of police use of excessive force. We predicted that those with stronger retributive attitudes would perceive that the police use excessive force less often, and we found that as attitudes became more retributive, perceptions of excessive force decreased.

A second important finding is the relationship between perceptions of excessive force and political ideology. The more liberal one's political ideology, the more often they believed that the police used too much force. Conversely, as political ideology became more conservative, respondents perceived that the police used excessive force less often. This aligns with prior findings that political attitudes influence views of police use of force.[6] In fact, a key finding of the study is that political ideology is the greatest predictor of perceptions of the frequency of police use of excessive force.

Findings concerning racial perspectives were not surprising. We found that respondents belonging to non-white race and ethnic groups perceive that the police use excessive force more frequently than white respondents, which supports the work of Drakulich et al., as they suggest, such variations in perceptions may be due to prior experiences with the police or racial resentment.[7] Other demographics such as sex, sexual orientation, income, and education were all predictors of perceptions of police excessive force, consistent with the findings of prior work.[5,6,7]

While previous scholars have considered many factors supporting police use of force, this study uniquely assesses how retributive attitudes influence perceptions of how often the police use too much force. Considering this survey was conducted after protests surrounding the police-involved deaths of George Floyd and Breonna Taylor, understanding contemporary perceptions of police excessive force is essential. Utilizing retributive attitudes provides additional insight into public perceptions of excessive force.

Despite this study's contribution, it does have limitations. First, utilizing a single measure, support for the death penalty, as a proxy for retributive attitudes. The ANES measures limited how retributive attitudes could be assessed. Future studies should consider an expanded measure of retributive attitudes. For example, other researchers have still employed death penalty support

as a proxy for retributive sentiments using multiple survey items, rather than a singular question.[17] As noted by Silver and Pickett, subsequent research on police force should consider the influence of perceptions of other criminal justice attitudes, aside from the death penalty, such as procedural justice, racial bias, confidence in the criminal justice system, and other police-specific practices.[6]

Another limitation of this study is the use of secondary data and relying on data from survey questions not specific to this study. Therefore, some of the indicators are vague. For example, respondents were asked, "How often do you think police officers use more force than necessary?" This question does not specify a scenario, such as during arrests or investigatory stops. Neither does it account for race-based disparities in the use of excessive force.

Future studies exploring police excessive force should include questions with context.[26,27,28] Distinguishing between the use of force and excessive force and accounting for situational factors may provide more nuanced outcomes. Also, employing qualitative methods to assess perceptions of police force and retributive attitudes would provide richer insight into individual perspectives and how they view the relationship between these concepts.

CONCLUSION

Police use of excessive force is one of the primary issues in current calls for police reform in the US. Assessing public perceptions of excessive police force allows for a more comprehensive understanding of society's view of police force. For public concern about police use of force and excessive force, research indicates that it is infrequent.[1,2] However, evidence demonstrates that these perceptions are greatly influenced by race and political ideology.[4,5,6,7] Disentangling perceptions from reality is critical to improving community-police relationships and developing evidence-based police reform.

REFERENCES

1. Eith, C., & Durose, M. R. (2011) *Contacts between police and the public, 2008*. Bureau of Justice Programs.
2. Kyprianides, A., Yesberg, J. A., Milani, J., Bradford, B., Quinton, P., & Oliver, C. (2021) Perceptions of police use of force: The importance of trust. *Policing, 44*(1), 175-190. *https://doi.org/10.1108/PIJPSM-07-2020-0111*
3. Alpert, G., & Dunham, R. (2004) *Understanding police use of force: Officers, suspects, and reciprocity*. Cambridge University Press. *https://doi.org/10.1017/CBO9780511499449*
4. Gerber, M. M., & Jackson, J. (2017) Justifying violence: Legitimacy, ideology and public support for police use of force. *Psychology, Crime & Law, 23*(1), 79-95. *https://doi.org/10.1080/1068316x.2016.1220556*
5. Mourtgos, S. M., & Adams, I. T. (2020) Assessing public perceptions of police use-of-force: Legal reasonableness and community standards. *Justice Quarterly, 37*(5), 869-899. *https://doi.org/10.1080/07418825.2019.1679864*
6. Silver, J. R., & Pickett, J. T. (2015) Toward a better understanding of politicized policing attitudes: Conflicted conservatism and support for police use of force. *Criminology, 53*(4), 650-676. *https://doi.org/10.1111/1745-9125.12092*
7. Drakulich, K., Robles, J., Rodriguez-Whitney, E., & Pereira, C. (2023) Who believes that the police use excessive force? Centering racism in research on perceptions of the police. *Journal of Research in Crime and Delinquency, 60*(1), 112-164. *https://doi.org/10.1177/00224278221120781*
8. Hollis, M. E. (2018) Measurement issues in police use of force: A state-of-the-art review. *Policing, 41*(6), 844-858. *https://doi.org/10.1108/PIJPSM-11-2017-0137*
9. Sunshine, J. & Tyler, T. R. (2003) The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review, 37*(3), 513-547. *https://doi.org/10.1111/1540-5893.3703002*
10. White, C., Hogan, M., Shelley, T., & N, P. U. (2018) The influence of procedural justice on citizen satisfaction with state law enforcement. *Policing, 41*(6), 687-703. *https://doi.org/10.1108/PIJPSM-02-2017-0026*
11. Bryant-Davis, T., Adams, T., Alejandre, A., & Gray, A. A. (2017) The trauma lens of police violence against racial and ethnic minorities. *Journal of Social Issues, 73*(4), 852-871. *https://doi.org/10.1111/josi.12251*
12. Bedau, H. A. (1978) Retribution and the theory of punishment. *Journal of Philosophy. 75*(11), 601-620.
13. Boersema, D. (2011) Retributive Justice. In: Chatterjee, D.K. (eds) Encyclopedia of Global Justice. Springer, Dordrecht. *https://doi.org/10.1007/978-1-4020-9160-5_381*
14. Bennett, C. (2014) Retributivist Theories. In: Bruinsma, G., Weisburd, D. (eds) *Encyclopedia of Criminology and Criminal Justice*. Springer. *https://doi-org.proxy.lib.utc.edu/10.1007/978-1-4614-5690-2_603*
15. Kelley, J., & Braithwaite, J. (1990) Public opinion and the death penalty in Australia. *Justice Quarterly, 7*, 529-563. *https://doi.org/10.1080/07418829000090721*
16. Thomas, C., & Foster, S. (1975) A sociological perspective on public support for capital punishment. *American Journal of Orthopsychiatry. 45*, 641-657. *https://doi.org/10.1111/j.1939-0025.1975.tb01192.x*
17. Liberman, Peter. 2013. Retributive support for international punishment and torture. *Journal of Conflict Resolution, 57*. 285–306. *https://doi.org/10.1177/0022002712445970*
18. Fitzgerald, R., & Ellsworth, P. C. (1984) Due process vs. crime control: Death qualification and jury attitudes. *Law and Human Behavior, 8*(1-2), 31-51. *https://doi.org/10.1007/BF01044350*

19. Tyler, T. R., & Weber, R. (1982) Support for the death penalty; instrumental response to crime, or symbolic attitude? *Law and Society Review, 17*, 21-45. *https://doi.org/10.2307/3053531*

20. Exum, J. J. (2015) The death penalty on the streets: What the eighth amendment can teach about regulating police use of force. *Missouri Law Review, 80*(4), 997-1010.

21. Cullen, F. T., Cao, L., Frank, J., Langworthy, R. H., Browning, S. L., Kopache, R., & Stevenson, T. J. (1996) "Stop or I'll shoot": Racial differences in support for police use of deadly force. *American Behavioral Scientist, 39*, 449–60.

22. Johnson, D., & Kuhns, J. B. (2009) Striking out: Race and support for police use of force. *Justice Quarterly, 26,* 592–623. *https://doi.org/10.1080/07418820802427825*

23. American National Election Studies. (2021) ANES 2020 Time Series Study Full Release [dataset and documentation]. July 19, 2020 version. *www.electionstudies.org*

24. Dimcock, M. (2019) Defining generations: Where Millenials end and Generation Z begins. Retrieved from: *https://www.pewresearch.org/short-reads/2019/01/17/where-millennials-end-and-generation-z-begins/*

25. Crawford, C. & Burns, R. (2008) Police use of force: Assessing the impact of time and space. *Policing & Society, 18*(3), 322-335. *https://doi.org/10.1080/10439460802292926*

26. Hollis, M. E. (2018) Measurement issues in police use of force: A state-of-the-art review. [Measurement issues in police use of force] Policing, 41(6), 844-858. *https://doi.org/10.1108/PIJPSM-11-2017-0137*

27. Thompson, B. L. & Lee, J. D. (2004) Who cares if police become violent? Explaining approval of police use of force using a national sample. *Sociological Inquiry, 74*(3), 381-410. *https://doi.org/10.1111/j.1475-682X.2004.00097.x*

ABOUT STUDENT AUTHOR

Amelia Collins graduated in May 2023 from the University of Tennessee at Chattanooga, in Chattanooga, Tennessee, where she was an Honor's College student and a double major in Criminal Justice and Psychology. Amelia is currently a first-year law student at the Samford University Cumberland School of Law.

PRESS SUMMARY

The purpose of this study was to examine the relationship between perceptions of police use of excessive force and retributive attitudes along with demographics. Findings indicated that respondents' perceptions of the frequency of police use of excessive force depend on their retributive attitudes. The more retributive one's attitude, the less often they perceived the police to use too much force. Similarly, the more conservative one's political ideology, the less frequently they perceived police used excessive force.

College Canines: Investigating the Behavioral and Physiological Impacts of Various College-Housing Environments on Companion Dogs

Kaitlyn Willgohs[*1], Jenna Williams[1], Isabella Crisostomo[2], Katherine Keck[1], Crystal Young-Erdos[2], & Lauren Highfill[1]*

[1]Psychology and Animal Studies, Eckerd College, St. Petersburg, FL
[2]Chemistry and Biochemistry, Eckerd College, St. Petersburg, FL

https://doi.org/10.33697/ajur.2023.090

*Student: kaitlyn.willgohs@gmail.com**
Mentors: highfile@eckerd.edu, youngecl@eckerd.edu

ABSTRACT
Companion animals are becoming a more familiar sight on college campuses, and they are often viewed as an essential element of wellness by students and institutions of higher education. While previous studies have investigated the behavioral and physiological impacts of bringing a pet to campus on the owners, impacts on the pets themselves have yet to be explored. Previous studies do suggest, however, that when dogs are left alone, they display more anxiety-related behaviors such as barking, destruction, lip-licking, body shaking, and higher levels of alertness. The present study investigated the difference in anxiety-related behaviors between on-campus dwelling dogs (n = 18) and off-campus dwelling dogs (n = 12) when exposed to a novel environment, and the physiological baseline of the dogs. Specifically, a saliva sample was collected from each dog before they were placed into a novel room for three minutes and their behavior was coded. Overall, there were no significant differences found between the two groups in either the anxiety-related behaviors observed or salivary cortisol levels. The implications of our findings for campus dogs will be discussed.

KEYWORDS
Companion Animals; Dogs; Behavior; Cortisol; Higher Education; Dog Welfare; Service Animals; Animal-Assisted Interventions; Student Mental Health

INTRODUCTION

Companion animals accompanying their owners when they go off to college encompasses a relatively new human-animal dynamic. Some colleges and universities are strict—only allowing service animals or emotional support animals—while other institutions offer entirely "pet-friendly" campuses. Regardless, the sight of companion animals on college campuses is becoming more prevalent.[1–2]

College can be a time of distress for students. Between transitioning from home to college, dealing with academic expectations, and managing new social relationships, students often experience anxiety, depression, and loneliness.[1–4] To mitigate these challenges, it is estimated that 62% of universities in the United States implement animal-visitation programs, such as visits from therapy dogs during exam weeks.[2,5] These programs are impactful, as previous literature suggests that human-animal interactions can have both physiological and psychological impacts on an individual, such as decreased blood pressure,[6–10] increased engagement in physical activity,[10–12] reductions in medication,[10, 13] greater feelings of happiness,[10, 14] and an improved mood.[10, 15]

Some colleges have gone as far as allowing students to bring their family pets to live with them on campus.[2, 16] However, bringing a family companion animal to live on a college campus introduces many changes for the animal such as living in a smaller space, being exposed to many unfamiliar people and animals, being left alone for long periods of time, and having to adjust to a new lifestyle and routine. Highfill and Goodman-Wilson (2017) found that student-pet owners, on average, reported a low investment of both time and money in the care of their pet.[17] Specifically, students reported spending an average of only four hours per day with their dogs or cats, which may render a more stressful living condition for animals than life in a family home.[17] While other studies have yet to investigate the impacts of living in a dorm room on dogs, researchers have examined the effects of other types of housing environments on stress levels in dogs. For example, Beerda et al. (1999) investigated stress levels in fifteen beagles living either in group housing or individually.[18] Overall, when living in housing that restricted their space and social interactions, dogs displayed more stress-related (*e.g.*, autogrooming, eating feces, vocalizing) and aggressive behaviors.[18]

Another potential issue of dogs living in a dorm setting is being left alone for extended periods of time while students are in class and at extracurricular activities. Previous studies have indicated that when left alone, dogs display more anxiety-related behaviors

such as barking,[19–21] destruction,[12] lip-licking, body shaking and higher levels of alertness.[19, 21] Frank et al. (2007) conducted an exploratory study investigating the behavior of puppies when left alone.[19] They found that many of the behaviors that were exhibited by the puppies were compatible with anxiety or fear, such as vocalizing, yawning, scratching, licking their lips, and a heightened sense of awareness.[19, 22]

In similar research, Rehn and Keeling (2011) investigated how dogs are affected when their owners leave them home alone during the workday.[21] In this study, twelve dogs were tested across three conditions: being left alone for thirty minutes, being left alone for two hours, and being left alone for four hours. Upon return of their owners, the dogs who were left alone for two and four hours demonstrated significantly more attentive behaviors and physical activity than the dogs who were left for half an hour. There was no difference in the number of interactions initiated by the owners, suggesting that the dogs' post-separation behaviors were dependent on the time differences rather than on owner behaviors. Overall, the researchers concluded that dogs could perceive and be affected by the length of time they are left alone, but researchers did not find any evidence to suggest that the welfare of dogs is reduced during separation.

In addition to behavioral observations, measuring cortisol concentrations is an increasingly common method for physiologically detecting dog stress and welfare.[23–24] Cortisol is a glucocorticoid hormone produced by the hypothalamic-pituitary-adrenal (HPA) stress response pathway. Therefore, this hormone has been used to investigate the stress of dogs in various environments such as shelters,[24–26] kennels/pet hotels,[27–28] laboratories,[29–31] veterinary clinics,[32–33] and the worksites of military, therapy, and other working dogs.[34–37] Detecting cortisol levels in non-invasive samples such as saliva has therefore become more widely used, as it is important that the methods used for measuring the physiological stress response in dogs does not cause additional stress.[23] Salivary cortisol concentrations correlate well with plasma levels and this non-invasive sample collection is generally well tolerated by dogs.[38]

Although companion animals are becoming an increasingly familiar sight on college campuses, previous studies have only investigated the effects of bringing a pet to college on pet owners. Therefore, the goal of the present study was to investigate the impacts of this situation on dogs by examining how on-campus and off-campus dwelling dogs respond to being left alone. We chose these two populations of dogs because they both live with college students, but they differ in their living spaces. We also measured if there were physiological differences in stress between the two groups of dogs by comparing baseline salivary cortisol levels. Based on previous literature, we hypothesized that when left alone, on-campus dwelling dogs would exhibit more anxiety-related behaviors than off-campus dwelling dogs, given that dogs who live on campus live in much smaller environments. Similarly, based on previous literature, we hypothesized that on-campus dwelling dogs would exhibit higher levels of baseline cortisol since they reside in smaller environments, suggesting that their increased stress could be detected both behaviorally and physiologically.

METHODS AND PROCEDURES

Subjects

This study was conducted at a small college in the southeastern United States and was approved by the home institution's IACUC. Thirty dogs (see **Table 1**) of varying breeds and sizes participated in the study (average age = 4.33 years; average length of ownership = 3.30 years; average length of time on-campus dwelling dogs have lived on campus = 2.33 semesters). In accordance with the previously reported effects of sex, neuter status, and age on salivary cortisol levels,[39] only dogs who were neutered or spayed and older than 6 months were selected for the study.

Name	Breed	Living	Age (yrs)	Length of Ownership (yrs)	Semesters Living on Campus	Analysis Type
Abby	Maltese Poodle	on	3	3	4	B; P
Apollo	Golden Retriever	off	0.75	0.75	n/a	B; P
Augie	American Eskimo Mix	on	1.5	1.5	2	B; P
Bean	Deer Head Chihuahua	on	12	2	3	B; P
Bosco	Black Lab/Rottweiler Mix	on	2	0.75	2	B; P
Callie	Mini Australian Shepherd	off	4	4	n/a	B
Chance	Border Collie	on	4	4	1	B; P
Cleo	Lab Mix	off	1.5	1	n/a	B; P
Dakota	Golden Retriever	on	5	5	2	B; P
Einstein	Pembroke Welsh Corgi	off	8	0.75	n/a	B; P
Ezra Grey	Australian Shepherd	on	1	1	1	B; P
Guppy	Cocker Spaniel	off	1.5	0.5	n/a	B
Ivy	Akita Inu	off	3	2.5	n/a	B; P
Jax	Jack Russel/Chihuahua Mix	off	3	3	n/a	B
Kona	Chihuahua Mix	on	5	2	3	B
Lexi	Lab Mix	on	12	12	1	B
Lila	Poodle Mix	on	5	5	6	B
Louie	Lab/Shepherd Mix	on	1	1	3	B; P
Luna	Pitbull Mix	on	2	2	4	B
Luna	Labrador Retriever	on	4	4	2	B; P
Luna	Golden Retriever	on	3	3	2	B; P
Molly	Husky Mix	on	8	7	1	B; P
Mookie	Pitbull/Bulldog Mix	on	2	0.5	1	B; P
Neo	Giant Schnauzer	off	9	9	n/a	B; P
Niki	Australian Shepherd	off	12	11	n/a	B; P
Padfoot	Havanese	off	3.5	3	n/a	B; P
Peanut	Chihuahua/Shih Tzu Mix	off	1	0.75	n/a	B; P
Rocky Balboa	Shih Tzu	on	5	5	2	B
Silas	Corgi Mix	on	3	1	2	B; P
Thea	Lab/Pitbull Mix	off	4	3	n/a	B; P

Table 1. Breed, living location, age, length of ownership, number of semesters living on campus, and the analysis of the subject (behavioral = B; physiological = P).

Participants were recruited via email and all dogs were owned by undergraduate students or recent alumni who were currently in graduate school and were offered either extra credit in their psychology courses or a $5 gift card if they participated in the study. Eighteen of the dogs lived on campus in "pet-friendly" dormitories. The remaining twelve dogs lived off campus with their owners.

The college has several styles of dormitories. Traditional housing dorms are approximately 3.4 m x 4.9 m, or 18.6 sq. meters and are double-occupancy bedrooms. All of the dorm room housing is much smaller than popular off-campus apartments, which range in size from 48.8 sq. meters (one bedroom, one bath) to 103.8 sq. meters (two bedrooms, two baths).

Testing Room

Testing took place in a small, carpeted lab room (3 m x 2.4 m). Owners were instructed to bring one of their dog's favorite items from home, such as a dog bed, a favorite blanket, or a toy, to leave in the room with their dog. Water was provided in a bowl.

The Outward Look of Fear/Anxiety/Stress	
Furrowed Brow	When the area above the dog's eyes shows tension and there is some wrinkling.
Panting: Tongue in Mouth	When the dog pants but the tongue doesn't protrude further than the dog's lower incisors.
Panting: Wide	When the dog pants and his mouth is open wider than seems necessary.
Penis Crowning	When the tip of the dog's penis sticks out of its sheath.
Spine Straight	When the dog's back remains in a straight line.
Whale Eye	When the whites of the dog's eyes show.
Information Gathering Signals	
Air-Sniffing (or, Leaning Forward Sniffing)	When the dog twitches and flares his nostrils, raising his nose upward or outward slightly, moves his nose from side to side. The dog's feet and body remain stationary.
Blinking	When the dog blinks his eyes less than every two seconds or longer (<2 seconds).
Licking: Nose	When the dog's tongue comes out the front of his mouth and completely covers his nose, so that for a brief moment you can see the underside of the dog's tongue. The tongue comes directly out and goes back in from the front of the dog's mouth.
Licking: Sideways	When the dog's tongue comes out the front of his mouth, covers the nose only partially and then moves down the side of the dog's mouth and retreats back in from the back of the lips.
Sniffing	When the dog sniffs the ground, furniture, people, objects for more than two seconds without disconnecting from scent, without chewing or eating anything (>2 seconds).
Action or Movement-Related Behaviors	
Hypervigilance	When the dog keeps in constant motion, or moves rapidly, and is constantly alert and aroused to his environment.
Hyperexcitability	When the dog's state of arousal is out of sync with what is going on in the environment—i.e. the dog is aroused when the situation is relatively stable. The dog will most often be panting quickly, eyes will be round and open, pupils dilated, brow furrowed, tail wagging.
Jumping	When both of the dog's hind legs leave the ground with front paws up.
Lack of Movement	When the dog's wakeful state keeps him standing, sitting or lying down, hardly moving his feet, his spine is usually straight, and there is little movement out of any part of his body (>3 seconds).
Stereotypic Behaviors	
Circling	When the dog traces a circular motion, in one direction, over and over again.
Pacing or Route-Tracing	When the dog traces a path side to side, walking a certain distance in one direction, and then the dog walks the same distance in the other direction, repeatedly. Each change in direction is preceded by a signature head-flick.
Vocalizations	
Bark Spells	Sharp loud vocalization. Number of seconds between barking spells (>1 second).
Whine / Cry	Long, high pitched vocalization. Measured frequency of whining / crying spells with at least 2 seconds separating each spell.

Table 2. Operational definitions of anxious-type behavior variables in dogs.

Behavioral Procedure

The dog and the owner experienced a brief acclimation period of two minutes, where the dog was allowed to sniff around the lab room and become comfortable with surroundings. Next, saliva was collected from the dog's mouth (see below for procedure).

Then, the owner was asked to do their normal routine of leaving the dog (*e.g.*, saying good-bye, petting). The dog was left alone in the lab room for three minutes. During the separation period, the owner was asked to step outside of the building, so that their scent was no longer in the vicinity of the lab room. The dog was observed by researchers through a one-way mirror. To reduce blind spots, the dog was filmed from two-points of view: (1) a camera on a table in the lab room, and (2) through the mirror. After 3 minutes, the owner returned to the room and greeted the dog. Videos were analyzed later and coded (inter-rater reliability = 91%) using an ethogram adapted from Sternberg (2007) (see **Table 2**).[40]

Physiological Procedure

All saliva samples were collected from 8:00 am-12:00 pm since cortisol levels are known to vary significantly throughout a 24-hour day.[39] To validate this four-hour window, mean salivary cortisol concentrations for on-campus dogs collected from 8:00 am-10:00 am (0.2564 μg/dL) versus 10:00 am-12:00 pm (0.1915 μg/dL) were compared and no significant difference was observed ($t(11) = -1.01, p = .17$). Dog owners were instructed to ensure that dogs did not eat one hour prior to sampling and did not drink 10 minutes before collection to avoid sample dilution.

Cortisol samples were only collected prior to behavioral analyses, as the goal of this measurement was to compare baseline cortisol levels, rather than to analyze any differences before and after the separation. Samples were collected using a Salimetrics SalivaBio Children's Swab by placing it on the side of the dog's mouth for ~30 seconds. Owners were asked to gently restrain their dog if necessary.

To facilitate salivation, the dog was presented with a small dog treat, which was given after sufficient saliva was collected (at least 0.5 g). Any samples that had visible blood contamination were discarded. The saliva was immediately frozen at -20 °C and later analyzed for cortisol using a Salimetrics Salivary Cortisol Elisa Kit according to the manufacturer's protocol. SPSS Statistics Version 26 was used to run independent-samples *t*-tests for both behavioral and physiological data.

RESULTS

Behavioral Results

A series of independent-samples *t*-tests were conducted to determine whether there were significant differences among observed anxiety-related behaviors between dogs living on-campus *vs.* off-campus (**Table 3**). Overall, no significant differences were found for any of the behaviors ($p > 0.05$).

Behavior	Living	Mean	Standard Deviation	t	df	p
Total # of Dog Anxious Behaviors	on off	23.94 24.00	14.95 13.78	-.010	28	.99
Sum of Outward Look of Fear/Anxiety/Stress	on off	6.97 6.42	6.14 5.58	.252	28	.99
Sum of Information Gathering Signals	on off	5.86 6.67	3.83 4.64	-.519	28	.61
Sum of Action or Movement-Related Behaviors	on off	6.14 6.50	5.73 5.04	-.177	28	.86
Sum of Stereotypic Behavior	on off	3.81 4.25	5.06 3.77	-.260	28	.80
Bark Spells	on off	3.75 2.67	4.95 4.52	.608	28	.55
Whine / Cry	on off	5.11 6.83	5.06 6.51	-.815	28	.42

Table 3. Independent samples t-tests of selected observed anxiety-related behaviors between dogs living on-campus vs. off-campus.

Physiological Results

While we were able to analyze the behaviors of all 30 dogs, analyzable saliva samples were collected from 22 of the dogs, 13 of those being on-campus dwelling and nine off-campus dwelling. All saliva samples had cortisol concentrations in the range of those commonly reported for domestic canines.[39] Even though the mean salivary cortisol concentration was higher for dogs living off-campus in comparison to those living on-campus (0.2946 μg/dL vs. 0.2115 μg/dL, respectively; (see **Figure 1**)), the difference was not statistically significant ($t(20) = -1.46, p = .08$;). The salivary cortisol levels of on-campus and off-campus dogs collected

in the first part of our sampling window (8:00am-10:00am) were further compared to ensure diurnal fluctuations were not masking any differences. No significant difference was found (on-campus mean: 0.1915 μg/dL; off-campus mean: 0.2839 μg/dL; $t(15) = -1.54$, $p = .07$).

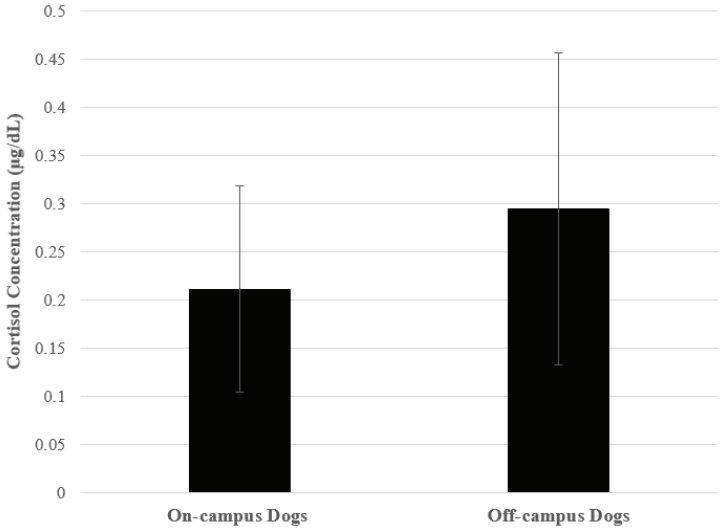


Figure 1. Comparison of salivary cortisol concentrations for on-campus ($n = 13$) vs. off-campus dogs ($n = 9$).

DISCUSSION

Based on previous research, we hypothesized that when left alone in the lab, on-campus dwelling dogs would exhibit more anxiety-related behaviors than off-campus dwelling dogs, given that dogs who live on-campus live in much smaller environments. Similarly, we hypothesized that on-campus dwelling dogs would exhibit higher levels of cortisol, suggesting that they were more stressed. Overall, our results indicate that there were no significant differences in cortisol levels and that there were no significant behavioral differences between the on-campus dwelling dogs and off-campus dwelling dogs when left alone. These results do not support our hypotheses; however, these results are positive in that they suggest that dogs living in dorm rooms on college campuses do not experience significantly different welfare conditions than dogs living in larger complexes off campus.

A confound to our research could be the duration of time that a dog has lived in their given environment. Rooney et al. (2007) suggested that dogs exhibit more anxiety-related behaviors and have higher levels of cortisol when living in a newer environment.[41] This is important to note since college students have the tendency to live in a dorm for nine months at a time before returning home for the summer. In this study, the average length of time that on-campus dwelling dogs lived on campus was two-and-one-third semesters. The average length of time that off-campus dwelling dogs lived in their respective environments was not obtained. Future studies should control for the duration that an animal has lived in their environment. Additionally, a longitudinal study could be conducted to compare the number of anxiety-related behaviors and the cortisol levels of dogs before they come to campus, immediately after they first come to campus, during their time on campus, and then again right before their owner graduates. An additional facet of this could include another non-invasive approach of analyzing the cortisol levels of hair over similar timelines to correlate the results with salivary concentrations.[41] Combined, this would provide more information on the impact of living on campus on dogs, as previous studies have suggested that dogs who are habituated to their environment may exhibit less anxiety-related behaviors and have lower levels of cortisol.[42]

Even though our results indicated no behavioral or physiological differences between on-campus and off-campus dwelling dogs, it is important to consider the mental stimulation of the dogs owned by college students. Indeed, previous studies have shown that increased human interaction has led to lower cortisol levels and improved scores on behavior tests in dogs residing in shelters.[25, 26, 43] Therefore, future studies could include a third subject group of off-campus dogs who are not owned by college students to explore possible differences in owner-lifestyle and dog welfare experiences. Additionally, to investigate the behavioral differences between on-campus and off-campus dwelling dogs when left alone in the home, future studies could utilize citizen science by asking owners to set up hidden cameras that film their dogs when left alone. Furthermore, given the limited sample size of our study—a challenge not new to comparative cognition[44–46])—we encourage other researchers to replicate our study to learn more about the impacts of various housing environments on companion dogs.

This study is among the first to explore the behavior and physiology of dogs living on college campuses. Since dogs are becoming an increasingly familiar sight on college campuses, it is of utmost importance to continue to research all aspects of welfare in order to ensure our furry friends are living fulfilling lives as their owners get their diplomas.

ACKNOWLEDGEMENTS

We extend our gratitude to the Ford Apprentice Scholars Program at Eckerd College, which provided financial support to purchase the materials necessary to collect cortisol samples and support to offer gift cards to participants. We thank the faculty at Eckerd College who offered extra credit to their students for participating, and we thank the students at Eckerd College who participated. We report no conflicts of interest or outside financial support for this research.

REFERENCES

1. Adams, A. C., Sharkin, B. S., & Bottinelli, J. J. (2017). The role of pets in the lives of college students: Implications for college counselors. *Journal of College Student Psychotherapy*, 31(4), 306–324. *https://doi.org/10.1080/87568225.2017.1299601*

2. Bruneau, L., & Johnson, A. (2021). Bringing Postsecondary Students Together With Dogs: Dog Welfare, Health, Safety and Liability Considerations. In M.R. Jalongo (Eds), *The Canine-Campus Connection: Roles for Dogs in the Lives of College Students* (pp. 57–96). Purdue University Press. *https://doi.org/10.2307/j.ctv16pn3th*

3. Adamle, K. N., Riley, T. A., & Carlson, T. (2009). Evaluating college student interest in pet therapy. *Journal of American College Health*, 57(5), 545–548. *https://doi.org/10.3200/jach.57.5.545–548*

4. Binfet, J. T. (2017). The effects of group-administered canine therapy on university students' wellbeing: A randomized controlled trial. *Anthrozoös*, 30(3), 397–414. *https://doi.org/10.1080/08927936.2017.1335097*

5. Haggerty, J. M., & Mueller, M. K. (2017). Animal-assisted stress reduction programs in higher education. *Innovative Higher Education*, 42(5), 379–389. *https://doi.org/10.1007/s10755-017-9392-0*

6. Allen, K., Shykoff, B. E., & Izzo, J. L. (2001). Pet ownership, but not ACE inhibitor therapy, blunts home blood pressure responses to mental stress. *Hypertension*, 38(4), 815–820. *https://doi.org/10.1161/hyp.38.4.815*

7. Baun, M. M., Bergstrom, N., Langston, N. F., & Thoma, L. (1984). Physiological effects of human/companion animal bonding. *Nursing Research*, 33(3), 126–129. *https://doi.org/10.1097/00006199-198405000-00002*

8. Friedmann, E., Katcher, A. H., Thomas, S. A., Lynch, J. J., & Messent, P. R. (1983). Social interaction and blood pressure. Influence of animal companions. *Journal of Nervous & Mental Disease*, *171*(8), 461–465. *https://doi.org/10.1097/00005053-198308000-00002*

9. Friedmann, E., Thomas, S. A., Cook, L. K., Tsai, C., & Picot, S. J. (2007). A friendly dog as potential moderator of cardiovascular response to speech in older hypertensives. *Anthrozoos*, *20*(1), 51–63. *https://doi.org/10.2752/089279307780216605*

10. Kist, S. E. (2009). *Correlates of pet-keeping in residence halls on college student adjustment at a small, private, Midwestern college*. University of Missouri-Columbia. *https://doi.org/10.32469/10355/6167*

11. Dembicki, D., & Anderson, J. (1996). Pet ownership may be a factor in improved health of the elderly. *Journal of Nutrition for the Elderly*, *15*(3), 15–31. *https://doi.org/10.1300/J052v15n03_02*

12. Serpell, J. (1991). Beneficial effects of pet ownership on some aspects of human health and behaviour. *Journal of the Royal Society of Medicine*, 84(12), 717–720. *https://doi.org/10.1177/014107689108401208*

13. Headey, B. (1999). Health benefits and health cost savings due to pets: Preliminary estimates from an Australian national survey. *Social Indicators Research*, *47*, 233–243. *https://doi.org/10.1023/A:1006892908532*

14. Ory, M. G., & Goldberg, E. L. (1983). Pet possession and life satisfaction in elderly women. In A. H. Katcher & A. M. Beck (Eds.), *New perspectives on our lives with companion animals* (pp. 303–317). Philadelphia University of Pennsylvania Press. *https://doi.org/10.1177/0164027583005003007*

15. Colby, P. M., & Sherman, A. (2002). Attachment styles impact on pet visitation effectiveness. *Anthrozoos*, *15*(2), 150–165. *https://doi.org/10.2752/089279302786992676*

16. Beach, L.A. (2019). More colleges are adopting pet friendly policies. *USA Today*. *https://www.usatoday.com/story/life/2019/05/19/animal-house/3705183002*

17. Highfill, L., & Goodman-Wilson, M. (2017). Pets on campus: best friend or bad decision?. *Human-Animal Interaction Bulletin*, 5(2), 1–13. *https://doi.org/10.1079/hai.2017.0012*

18. Beerda, B., Schilder, M. B., Van Hooff, J. A., De Vries, H. W., & Mol, J. A. (1999). Chronic stress in dogs subjected to social and spatial restrictions. I. Behavioral responses. *Physiology and Behavior, 66*(2), 233–242. *https://doi.org/10.1016/S0031-9384(98)00289-3*

19. Frank, D., Minero, M., Cannas, S., & Palestrini, C. (2007). Puppy behaviours when left home alone: a pilot study. *Applied Animal Behaviour Science*, 104(1–2), 61–70. *https://doi.org/10.1016/j.applanim.2006.05.003*

20. Miklosi, A., Polgárdi, R., Topál, J., & Csányi, V. (2000). Intentional behaviour in dog-human communication: an experimental analysis of "showing" behaviour in the dog. *Animal cognition*, 3(3), 159–166. *https://doi.org/10.1007/s100710000072*

21. Rehn, T., Keeling, L. J. (2011). The effect of time left alone at home on dog welfare. *Applied Animal Behaviour Science*, *129*(2–4), 129–135. *https://doi.org/10.1016/j.applanim.2010.11.015*

22. Overall, K. L., Dunham, A. E., & Frank, D. (2001). Frequency of nonspecific clinical signs in dogs with separation anxiety, thunderstorm phobia, and noise phobia, alone or in combination. *Journal of the American Veterinary Medical Association*, 219(4), 467–473. *https://doi.org/10.2460/javma.2001.219.467*

23. Chmelíková, E., Bolechová, P., Chaloupková, H., Svobodová, I., Jovičić, M., & Sedmíková, M. (2020). Salivary cortisol as a marker of acute stress in dogs: A review. *Domestic Animal Endocrinology*, *72*, 106428. *https://doi.org/10.1016/j.domaniend.2019.106428*

24. Hennessy, M. B. (2013). Using hypothalamic-pituitary-adrenal measures for assessing and reducing the stress of dogs in shelters: A review. *Applied Animal Behaviour Science, 149*(4), 1–12. *https://doi.org/10.1016/j.applanim.2013.09.004*

25. Coppola, C. L., Grandin, T., Enns, R. M. (2006). Human interaction and cortisol: Can human contact reduce stress for shelter dogs? *Physiology & Behavior*, *87*(3), 537–541. *https://doi.org/10.1016/j.physbeh.2005.12.001*

26. Menor-Campos, D. J., Molleda-Carbonell, J. M., & López-Rodríguez, R. (2011). Effects of exercise and human contact on animal welfare in a dog shelter. *Veterinary Record*, *169*(15), 388. *https://doi.org/10.1136/vr.d4757*

27. Grigg, E. K., Nibblett, B. M., Robinson, J. Q., & Smits, J. E. (2017). Evaluating pair versus solitary housing in kenneled domestic dogs (*Canis familiaris*) using behaviour and hair cortisol: A pilot study. *Veterinary Record Open*, *4*(1), 1–14. *https://doi.org/10.1136/vetreco-2016-000193*

28. Wojtaś, J., Garbiec, A., Karpiński, M., & Czyżowski, P. (2022). Dogs' Stay in a Pet Hotel–Salivary Cortisol Level and Adaptation to New Conditions. Journal of Applied Animal Welfare Science, 25(3), 208–213. *https://doi.org/10.1080/10888705.2020.1781631*

29. Beerda, B., Schilder, M. B., Van Hooff, J. A., De Vries, H. W., & Mol, J. A. (1998). Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. *Applied Animal Behaviour Science*, 58(3–4), 365–381. *https://doi.org/10.1016/s0168-1591(97)00145-7*

30. Spangenberg, E. M. F., Bjorklund, L., Dahlborn, K. (2006) Outdoor housing of laboratory dogs: Effects on activity, behavior and physiology. *Applied Animal Behaviour Science*, *98*(3–4), 260–276. *https://doi.org/10.1016/j.applanim.2005.09.004*

31. Stracke, J., Bert, B., Fink, H., & Böhner, J. (2011) Assessment of stress in laboratory beagle dogs constrained by a Pavlov sling. *ALTEX*, *28*(4), 317–325. *http://dx.doi.org/10.17169/refubium-21300*

32. Hekman, J. P., Karas, A. Z., & Dreschel, N. A. (2012). Salivary cortisol concentrations and behavior in a population of healthy dogs hospitalized for elective procedures. *Applied Animal Behaviour Science*, *141*(3–4), 149–157. *https://doi.org/10.1016/j.applanim.2012.08.007*

33. Siracusa, C., Manteca, X., Cerón, J., Martínez-Subiela, S., Cuenca, R., Lavín, S., Garcia, F., & Pastor, J. (2008). Preoperative stress response in dogs undergoing elective surgery: Variations in behavioural, neuroendocrine, immune and acute phase responses. *Animal Welfare*, *17*(3), 259–273. *https://doi.org/10.2460/javma.237.6.673*

34. Arcuri, G. B., Pantoja, M. H. A., Titto, C. G., & Martins, D. D. S. (2022). Preliminary analysis of reproductive, behavioral and physiological characteristics of military working dogs. *Animal Reproduction*, *19*(1), e20210092. *https://doi.org/10.1590/1984-3143-AR2021-0092*

35. Glenk, L. M., Kothgassner, O. D., Stetina, B. U., Palme, R., Kepplinger, B., & Baran, H. (2014). Salivary cortisol and behavior in therapy dogs during animal-assisted interventions: A pilot study. *Journal of Veterinary Behavior*, 9(3), 98–106. *https://doi.org/10.1016/j.jveb.2014.02.005*

36. Haverbeke, A., Diederich, C., Depiereux, E., & Giffroy, J. M. (2008). Cortisol and behavioral responses of working dogs to environmental challenges. *Physiology & Behavior*, *93*(1–2), 59–67. *https://doi.org/10.1016/j.physbeh.2007.07.014*

37. Ng, Z. Y., Pierce, B. J., Otto, C. M., Buechner-Maxwell, V. A., Siracusa, C., & Werre, S. R. (2014). The effect of dog–human interaction on cortisol and behavior in registered animal-assisted activity dogs. *Applied Animal Behaviour Science*, 159, 69–81. *https://doi.org/10.1016/j.applanim.2014.07.009*

38. Vincent, I. C., & Michell, A. R. (1992) Comparison of cortisol concentrations in saliva and plasma of dogs. *Research in Veterinary Science*, *53*(3), 342–345. *https://doi.org/10.1016/0034-5288(92)90137-Q*

39. Cobb, M.L., Iskandarani, K., Chinchilli, V. M., Dreschel, N. A. (2016). A systematic review and meta-analysis of salivary cortisol measurement in domestic canines. *Domestic Animal Endocrinology*, *57*, 31–42. *https://doi.org/10.1016/j.domaniend.2016.04.003*

40. Sternberg, S. (2007). *Fear ~ Stress Behaviors*. Dog Ethogram. *https://www.suesternberg.com/*

41. Bryan, H. M., Adams, A. G., Invik, R. M., Wynne-Edwards, K. E., & Smits, J. E. (2013). Hair as a meaningful measure of baseline cortisol levels over time in dogs. *Journal of the American Association for Laboratory Animal Science*, *52*(2), 189-196.

42. Rooney, N. J., Gaines, S. A., & Bradshaw, J. W. (2007). Behavioural and glucocorticoid responses of dogs (*Canis familiaris*) to kennelling: Investigating mitigation of stress by prior habituation. *Physiology & Behavior*, *92*(5), 847–854. *https://doi.org/10.1016/j.physbeh.2007.06.011*

43. Hennessy, M. B., Davis, H. N., Williams, M. T., Mellott, C., Douglas, C.W. (1997). Plasma cortisol levels of dogs at a county animal shelter. *Physiology & Behavior, 62*(3), 485–490. *https://doi.org/10.1016/S0031-9384(97)80328-9*

44. Agrillo, C., & Miletto Petrazzini, M. E. (2012). The importance of replication in comparative psychology: the lesson of elephant quantity judgments. *Frontiers in Psychology*, *3*, 181. *https://doi.org/10.3389/fpsyg.2012.00181*
45. Beran, M. J., Parrish, A. E., Perdue, B. M., & Washburn, D. A. (2014). Comparative cognition: Past, present, and future. *International Journal of Comparative Psychology. https://doi.org/10.46867/ijcp.2014.27.01.07*
46. Stevens, J. R. (2010). The challenges of understanding animal minds. *Frontiers in Psychology*, *1*, 203. *https://doi.org/10.3389/fpsyg.2010.00203*

ABOUT STUDENT AUTHORS

All students completed this research while undergraduate students at Eckerd College, and have gone on to continue their respective educations. Kaitlyn Willgohs graduated from Eckerd College in May 2021 and began pursuing her Ph.D. in cognitive and comparative psychology at the Graduate Center, City University of New York, in August 2023. Jenna Williams graduated from Eckerd College in May 2022 and began pursuing her M.Sc. in biological sciences from the University of Manitoba in August 2022. Isabella Crisostomo graduated from Eckerd College in 2020 and began pursuing her M.D. at the University of Colorado, Anschutz, in August 2023. Katherine Keck graduated from Eckerd College in 2020 and graduated with her M.Sc. in evolutionary and comparative psychology from the University of St. Andrews in 2021.

PRESS SUMMARY

Companion animals are becoming a more familiar sight on college campuses, and they are often viewed as an essential element of wellness by students and institutions of higher education. Though previous researchers have investigated the behavioral and physiological impacts of bringing a pet to campus on the owners, the present study is the first to investigate the impact on the pets themselves. Specifically, the present study investigated the difference in anxiety-related behaviors between on-campus dwelling dogs and off-campus dwelling dogs when exposed to a novel (new) environment, and the physiological baseline of the dogs. Overall, there were no significant differences found between the two groups in either the anxiety-related behaviors observed or salivary cortisol levels, suggesting that there were no behavioral or physiological differences between these two populations of dogs.

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