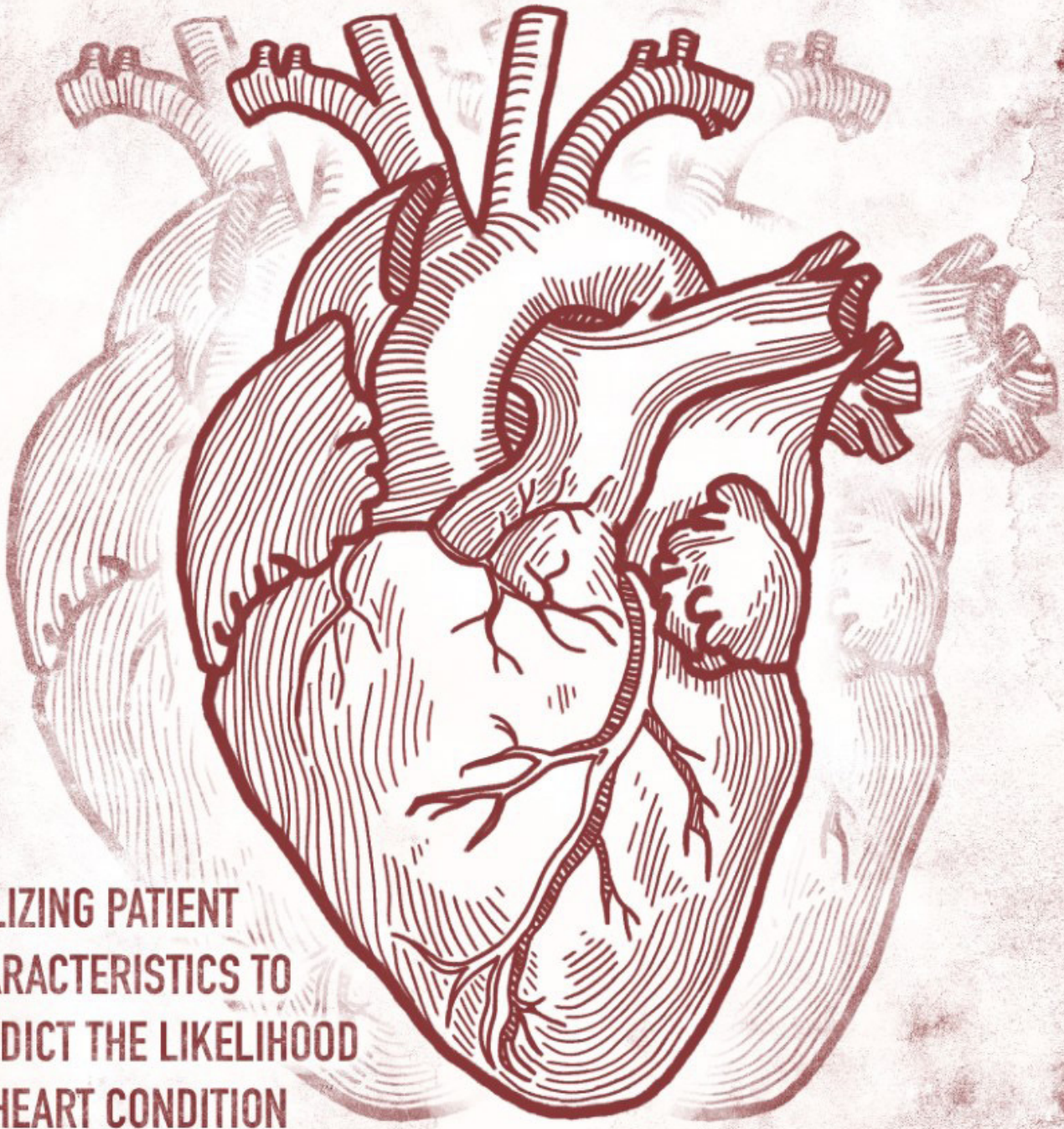
VERTICES

Duke's Undergraduate Research Journal

Spring 2022
Volume I, Issue I



UTILIZING PATIENT CHARACTERISTICS TO PREDICT THE LIKELIHOOD OF HEART CONDITION

Table of Contents

Letter from the Editor

Dear Reader,

I'm delighted to present the first issue of Vertices: Duke's Undergraduate Research Journal. The Spring 2022 semester signifies a huge step for our organization. It marks the transformation of Vertices from an article repository into a full-fledged journal. I am so grateful for the editors, authors, artistic and website designers, and Vertices members that made this next step possible.

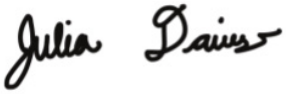
The covid-19 outbreak and remote classes hit the academic editing component of Vertices hard. Few articles were being reviewed and none were being published. But in Fall 2021, we were determined to turn this around. Thanks to a team of 13 dedicated peer reviewers, we published seven stellar pieces of undergraduate research on our website.

Going into Spring 2022 we wanted something more: to transform Vertices' academic publishing to fit the model of a rigorous academic undergraduate journal. To do this, we established an editorial calendar, with the ultimate goal of publishing two cohesive issues per year, Fall and Spring. In this inaugural issue, we expanded our pool of editorial reviewers to include students at Georgetown, as well as Duke faculty. For the first time in Vertices history, we were able to register each article with a Digital Object Identifier (DOI) to make the manuscripts easily accessible online. This semester we grew collectively as a student organization and as an academic research publication.

I'm honored to share the finished publication here. From all of the submissions this semester, we have carefully chosen and reviewed four strong pieces to showcase. The article featured on the cover of this issue is traditional scientific research, applying statistics to cardiology. Alongside this piece, we have a social science publication exploring attitudes toward male contraception—an increasingly relevant topic in our nation. Our third article brings psychological background and statistical analysis to a newly established peer mentorship program at Duke. Finally, our last article explores the relationship between video games and player emotion as gaming expands its influence worldwide.

In this issue, we provide a glimpse into the broad range of high quality research that students are engaged in at Duke. I hope you will join in my excitement as you read this collaborative, thought-provoking publication produced by members of our Vertices team.

Sincerely,



Julia Davis, Editor in Chief

Masculinity and Novel Male Contraceptives: Does Masculine Norm Conformity Influence Preference?

Connie Dean

Volume 1, Issue 1

Masculinity and Novel Male Contraceptives: Does Masculine Norm Conformity Influence Preference?

Connie Dean

Department of Psychology and Neuroscience, Duke University
https://doi.org/10.55894/dv1.11

Abstract

Background: The impact of masculine norm endorsement on decision making regarding male contraceptive methods is a bounteous, yet unexplored, area of research.
Aim: The aim of this paper is to understand how masculinity and attitudes towards sex impact male contraceptive usage and how preference could inform strategies to ultimately improve uptake and usage. Increased contraceptive use could have profound public health and financial impact given the potential to prevent unwanted pregnancies and promote equitable family planning.
Method: 103 male participants aged 18+ completed a self-report electronic survey assessing masculine norm endorsement, attitudes towards sexuality and likelihood-of-usage of three novel male contraceptives; namely Nestorone, MENT and RISUG. Masculine norm endorsement was measured using an adapted version of Parent and Moradi's Conformity to Masculine Norms Inventory-46, and the Sexual Opinion Survey was used to assess positive-negative attitudes toward sex. Participants were also asked which of the three contraceptive they would be most likely to use and why
Results: The data showed no significant correlation between either masculine norm endorsement or attitudes towards sexuality and likelihood-of-usage for novel male contraceptives. However, the sample did show a hierarchy of preference; Nesterone was the most preferred method, followed by RISUG, and finally the sample indicated a significant aversion to MENT. For each contraceptive, common themes for preference were easily identifiable. Those who preferred Nesterone cited 'Non-invasiveness' and 'Ease of use' as their reasoning. Those who preferred RISUG indicated the 'Convenience' and 'Longevity' of the method as the reason for preference. Those who preferred MENT, cited 'Non-daily use' and 'Less invasive [than RISUG]'. Those who preferred MENT seemed to fit a niche between RISUG and Nesterone: while they do not wish to partake in the daily application of Nesterone, they are also opposed to the procedure involved with RISUG.
Conclusions: The results of this study indicate that the sample showed preference with regards to the novel forms of male contraceptives. The reasons indicated by the sample regarding their preference pointed to specific driving factors, motivated by what the individual values. Namely: ease of use, longevity, and surgical procedure avoidance. Furthermore, this suggests the need for multiple methods of male contraceptives to satisfy a broader range of individuals. Thus, this study yields important data that may help to inform future contraceptive design and research to meet the needs and preferences of young adult males.

Keywords: contraceptives; masculinity; sexuality; conformity

Introduction

Contraception is the most effective way to avoid unplanned pregnancies (Sonfield et al., 2013)–which comprise over 50% of all pregnancies (Bearak et al., 2018). Whilst women account for 70% of overall contraceptive use, there remain only two forms of contraception for men: condoms and vasectomy. Male contraceptives would not only have a measurable impact on the unplanned pregnancy rate,

but would also allow for greater female reproductive freedom and an equal sharing of the reproductive burden (Dorman, 2018). The demand for this change is tangible, with men and women reporting interest in having novel male contraceptive options available (Dorman & Bishai, 2012). With numerous novel male contraceptives in various stages of pre-clinical and clinical trials, identifying ways in which to integrate these novel methods into the reproductive landscape is paramount to maximize uptake and usage. Two important predictors of male contraceptive uptake are attitudes towards sex and masculine norm conformity. More specifically, higher masculine endorsement and more negative attitudes towards sexuality can predict avoidance or reluctance to use contraceptives (Lohiya et al., 2014; Fisher, 1984). However, it is currently unknown whether these same attitudes will apply to novel contraceptive methods. The goal of this study is to better understand how masculinity and attitudes towards sex dictate novel male contraceptive usage and preference, in order to best inform research and marketing strategies and maximize uptake. Increased male contraceptive use will have profound public health implications by reducing unwanted pregnancies, promoting equitable family planning, and addressing reproductive health barriers among women.

Masculinity and Health

Gender plays a crucial role in health-related decisions and mortality. On average, American men tend to die 7 years earlier than American women (Courtenay, 1996), suffer from more chronic health conditions (Verbrugge & Wingard, 1987), and underuse health care services compared to their female counterparts (Banks, 2004). Conformity to masculinity presents itself as a contributing factor to these gender differences. Men who conform more rigidly to conventional masculine norms – such as the importance of self-reliance, power over women and heterosexual presentation - are less likely to seek professional medical help and more likely to experience adverse health consequences (Addis & Mahalik, 2003). In the realm of reproductive health, poor-health decisions include avoidance of contraceptives and consequent disease transmission or unplanned pregnancy, and thus are of importance to this current study.

Contraceptive choices and Masculine norms

Whilst endorsement of general masculine norms has been associated with avoidance of new contraceptives (Lohiya et al., 2014), several studies have sought to separate masculine norms into categories, in order to obtain a more comprehensive view of how specific norms predict contraceptive usage. The nine norms used in Parent and Moradi's Conformity to Masculine Norms Inventory 46 (CMNI-46) have been widely assessed in their relation to contraceptive use. When masculine norms are separated into these categories, some relate more strongly to certain contraceptive beliefs and decisions. For example, self-rated endorsement of Heterosexual Self-Presentation norms (concern with being perceived as heterosexual) has been linked to avoidance of condoms (Limmer, 2016), whilst endorsement of 'Power Over Women' (preference to have control over women and for men to have societal control) predicts men's rejection of vasectomy (Martinez & Ramos, 2018). A recent study assessing men's 'willingness-to-try' a novel male contraception called RISUG, found both Heterosexual Self-Presentation and Power Over Women were negative predictors of the participants willingness-to-try (Lacasse et.al, 2019). If male contraceptives are to be a success, one must consider the particular masculine norms that impact aversions to usage in order to guide marketing and research.

When men will seek help

There are, however, instances when men will seek help. This occurs when the help seeking will preserve a more prominent component of the individual's masculinity (O'Brian et al., 2005), such as ensuring the functionality of the penis, particularly in men who conform to the notion that demonstrations of sexual performance and power are expressed through sustained erections, prolonged sexual intercourse, and penetration (Khann et al., 2008). In instances such as this, men often seek out help free of stigma (O'Brian et al., 2005). If the specific norms that dictate help-seeking behaviours are isolated, they may act as a form of guidance to shape targeted messaging campaigns or in identifying specific side effects that would be considered unacceptable in relation to masculine

norms.

Attitudes towards sex and contraceptive usage

The Sexual Behavior Sequence model (Byrne, 1977; 1983) describes an individual's response to sexual stimuli on a continuum from negative (erotophobia) to positive (erotophilia) and suggests that one's emotional response to sexuality may influence attitudes toward contraception. The model hypothesizes that negative responses to sexuality predicts avoidance of contraception (Fisher, 1984); One study found that sexually active college men with higher erotophilia ratings were more likely to use condoms (Fisher, 1984). Given that individuals with negative feelings about sexuality and sex tend to avoid learning about or using contraceptives (Fisher, 1984), this makes attitudes towards sex/sexuality an important concept in the realm of novel male contraceptives, as a means by which to inform sex education and marketing strategies going forward.

Novel male contraceptives

Previous research has shown a demand for more male contraceptive options, with many men interested in using novel male contraceptives (Dorman & Bishai, 2012). Promisingly, there are numerous novel male contraceptives in various stages of pre-clinical and clinical trials.

Reversible Inhibition of Sperm Under Guidance (RISUG) is a styrene maleic anhydride (SMA) hydrogel polymer dissolved in dimethyl sulphoxide (DMSO) that coats the lumen wall of the vas deferens upon direct injection. The SMA polymer creates a positively charged and acidic environment, that damages sperm as they pass through the vas deferens. Thus, upon ejaculation, the sperm are sufficiently damaged that they are considered infertile (Khilwani et al., 2020). The injection of RISUG is considered a minor surgical procedure and has up to 10 years of efficacy in a given individual. RISUG is currently in Phase-III of clinical trials and has so far been seen to be both safe and effective (Khilwani et al., 2020; Lohiya et al., 2014).

Nestrone is a progestin; a synthetic version of the progestogen hormone. Nesterone is a transdermal gel often co-administered with transdermal testosterone to illicit contraceptive effects and is thought to do so by inhibiting both gonadotropin release from the pituitary gland, and local testosterone production in the testes (Illani et al 2012; Kumar et al 2017). When 8mg of the gel is applied to the shoulders once daily, this contraceptive reduces sperm count to ≤ 1 million/mL; which is considered infertility (Illani et al 2012). Nesterone is in Phase IIb trails (Anawalt, et al., 2019).

MENT (7α-methyl-19-nortestosterone) is a synthetic androgen; a steroid hormone that acts upon androgen receptors and regulates male sex characteristic development and expression (Von Eckardstein et al., 2003; Sundaram et al., 1993). MENT acts as a substitute for testosterone, and similarly to Nesterone, effects gonadotropin secretion. MENT is administered as a subdermal implant that is placed under the skin of the upper arm, lasting up to one year (Sundaram et al., 1993) and has undergone clinical trials to access efficacy (Von Eckardstein et al., 2003; Chao et al., 2014; Anderson & Baird, 2002; Sundaram et al., 1993).

The current study

This study aims to assess endorsement of specific masculine norms, attitudes towards sexuality, and how both impact preference regarding novel male contraceptives. The current research will utilize online self-report survey methodology to collect the necessary data. The three novel contraceptives chosen were Nesterone, MENT and RISUG. These three were chosen because they are the furthest along the clinical trial pathway and have thus far offered the most promising results.

It is hypothesized that high conformity to masculine norms will predict lower likelihood-of-usage of novel male contraceptives. Furthermore, men who highly endorse the 'self-reliance' norm as defined by the CMNI-46 will favor the only contraceptive that can be self-administered; Nesterone. Alternatively, those who highly endorse the 'Playboy' norm of the CMNI-46 will avoid MENT. MENT involves a subdermal implant that reduces volume of ejaculate during sexual intercourse. Previous literature has shown that the functionality of the penis is a valued masculine norm, particularly in men who believe that sexual performance is expressed through erections,

ejaculation, and penetration (Khann et al., 2008). Thus, this contraceptive impacting ejaculate volume may be seen as a threat to sexual performance. In terms of attitudes towards sexuality, it is hypothesized that positive attitudes towards sex (erotophilia) will predict higher average likelihood of usage of novel male contraceptives.

Method

Participants

The population of interest included individuals of reproductive age who produce ejaculate and have sexual interactions with partner(s) who could potentially become pregnant. Predominantly, these individuals were cisgender males of heterosexual or bisexual orientation; however, transgender persons and gay men also fell in this category. 220 volunteers gave informed consent to participate in this study via electronic survey, which was approved by the Duke University Institutional Review Board. Of the 220 initial responses, there were 102 responses that answered the survey to completion. This study examines male contraception exclusively and likelihood of usage for the target population. Cisgender women will not be the ones administering the male contraception and thus their ratings regarding likelihood-of-usage will not be useful for this research question. In order to identify the target population, screening questions were included prior to data collection. These assessed an individual's gender, whether they produce ejaculate, and whether they were sexually active with partner(s) who could get pregnant. If they did not meet these criteria, they did not proceed with the survey.

Measures

The following measures were administered and accessed through an electronic survey and can be found in the Appendix. The relationship between these variables - likelihood-of-usage the contraceptive methods, endorsement of masculine norms, and attitudes toward sex – can indicate driving factors for contraceptive preference. Alongside the qualitative reasons for preference, these relationships can inform strategies to improve uptake and usage of male contraception in the future. Strategies to expand male contraceptive use could have profound public health and financial impact given the potential to prevent unwanted pregnancies.

Conformity to Masculine Norms

To assess endorsement of masculine norms, an adapted version of Parent and Moradi's 2009 Conformity to Masculine Norms Inventory 46 (CMNI-46) was used. When deciding on which masculine norms to assess male contraceptive preference, it was important to consider the specifics of the chosen male contraceptives and the purpose of the study. As previously mentioned, some norms are more closely related to contraceptives, while some predict more general behavior. For example, endorsement of Primacy of Work and Winning norms in the CMNI may well predict men's lack of exercise (Garfield et al., 2008), while endorsement of Risk-Taking predicts alcohol problems (Iwamoto et al., 2011), conceptually it is unlikely that either scale would provide further insight into contraceptive decision behaviors. However, the recent Lacasse paper assessing masculine norms and male contraceptives RISUG indicated both Heterosexual Self-Presentation and Power Over Women were negative predictors (Lacasse et.al, 2019). Thus, the CMNI was adapted to the purpose of this study in that only certain subscales were administered. The original CMNI-46 consists of 46 questions pertaining to nine norms: Winning, Emotional Control, Risk-Taking, Violence, Playboy, Self-Reliance, Primacy of Work, Power over Women and Heterosexual Self-presentation. The norms included were: Playboy, Self-Reliance, Power over Women and Heterosexual Self-presentation. Based on a literature review, these norms consistently interacted with health decision making regarding reproductive health and contraceptives and thus were the most salient for the purpose of this study.

Attitudes towards sex and sexuality

To assess positive-negative attitudes toward sex, researchers employed the Sexual Opinion Survey (Rye et al, 2011), which has been widely used in studies pertaining to HIV education, STD prevention, and contraceptive use. Whilst the 'Playboy' measure in

the CMNI-46 does provide some insight into desire for multiple sexual partners, it does not explicitly indicate an individual's positive or negative attitudes toward sexuality. To measure erotophobia-erotophilia, the Sexual Opinion Survey (SOS) was chosen. The SOS is a 21-item scale assessing the individual's attitudes toward different sexual behaviors (e.g., masturbation, fantasizing). Of the 21 items, 11 assess erotophilia/ positive attitudes (e.g., "Masturbation is an enjoyable experience") and 10 assess /negative attitudes to (e.g., "I am not curious about exotic erotica"). Each item is rated on a 4- point scale (1 = strongly agree, 4 = strongly disagree; Rye et al, 2011; Macapagal et al., 2011).

Likelihood-of-usage of Novel MCM

To measure likelihood-of-usage a novel male contraceptive, the participants were given a short summary about the three contraceptives (RISUG, Nestorone®, MENT®) and then asked to rate their likelihood-of-usage on a Likert scale (1 = extremely likely, 5 = extremely unlikely). Second, they were asked to choose one of the three methods as a preferred method and indicate why.

Procedure

Data was collected online between October 9th, 2020 and February 12th, 2021 via an electronic survey distributed through an anonymous link shared independently and on social media. Informed consent was obtained at the beginning of the survey, as mandated by IRB protocol. Consent was obtained through selecting "yes" and "continue" in lieu of an online signature. The survey began with specific screening questions to ensure only the target population was included in analysis. These assessed whether individuals produced ejaculate, their gender and whether they were sexually active with a partner(s) who could get pregnant. If the participants were not screened out, they were assessed for their Likelihood-of-usage for RISUG, Nestorone®, MENT individually. They then choose one of the three as a preferred method and explained why they made that choice. Participants were then answered 19 questions from the adapted version of the CMNI-46, and 21 questions from the SOS. When the participants finished the survey, they were taken to a 'Thank you' screen and were notified of their completion.

Preparing Data for Analysis

Data analysis was conducted by the undergraduate researcher, with assistance from their graduate mentor. The data used for analysis were obtained through the output downloaded from Qualtrics® electronic survey software, in the form of a Microsoft Excel spreadsheet. Quantitative analysis was conducted in JASP.

Coding

For the purposes of coding responses to the one qualitative measure, the aforementioned Microsoft Excel spreadsheet was uploaded into the qualitative data analysis software program, Dedoose. Using Dedoose, the undergraduate researcher separated out responses for each of the three contraceptives. These individual responses were then categorized into varying code items depending on the common responses and themes for that contraceptive. The undergraduate researcher created a data dictionary in for each contraceptive, in correspondence to common themes in responses. These code categories were then shown to the graduate mentor for approval, and agreement. For RISUG each response was categorized into one of the following categories: Reversibility, Forgetful, Ease/no hassle/ convenience, Effective, Longevity, Invasive, One-time use, Ejaculate Reduction and Minimal side effects. For Nestorone each response was categorized into one of the following categories: Most control/ flexibility, Ease of use, Non-invasive, Least painful, Easy to stop, Non-surgical, Safety and No doctors. For MENT each response was categorized into one of the following categories: Less invasive, Non-daily use, Pain-free and Low maintenance. Upon both coders completing their coding, the result of the interrater reliability test included overall Pearson's correlation coefficient, Pearson's r for each of the codes included in the test and relative difference metrics for each code. Code applications were collected from coding charts in the "Analyze" page on the Dedoose website (Analyze > Code Charts > Code Application).

Results

Sample

Of the 220 individuals who clicked on the survey link, 215 signed the consent form. Three individuals were removed for being female and of the remaining participants three were Transgender, four were transfeminine, six were assigned male at birth and one was intersex, and the remainder were male. 13 participants were removed for not producing ejaculate and 30 were removed for not having sex with partners who can become pregnant. Of the remaining 169 participants, 103 finished the survey to completion. There was an attrition rate of 39.1%.

Internal Consistency of Variables

The first analysis run was an internal reliability test for each measure independently. For Conformity to Masculine norms, internal reliability was high with Cronbach' alpha (a) = 0.725, thus no questions were excluded from analysis. For the second measure of 'Likelihood of usage', Cronbach' alpha (a) = 0.705, and no questions were excluded from analysis. For the variable 'Attitudes toward sexuality' (operationalized using the Sexual opinion survey), internal reliability was run using all 21 questions and Cronbach' alpha (a) =0.708. Again, no changes were made.

General Descriptive Statistics

An Alpha level of $p<0.05$ was used for all statistical analyses. See Table 1a, 1b and 1c for full descriptive information.

Conformity to Masculine Norms

Descriptive plots displaying density showed a positive skew for conformity to masculine norm average across the sample. Lower ratings on the scale indicate lower conformity to masculine norms (See Figure 1). Furthermore, a one sample t-test showed participants exhibited a statistically significant difference from typical endorsement of masculine norms as evidenced by lower scores (M= [2.222], SD =[0.723]), when compared to the neutral 2.5 value on the 4 point Likert scale: $t(102)=-3.896$ and $p<.001$ with Cohen's d (CD) = 0.3845 (See Figure 2). The effect size for this analysis was found to be between a small and medium effect size as defined by Cohen's (1988) conventions for a small (d = .20), and medium (d=.50) effect size.

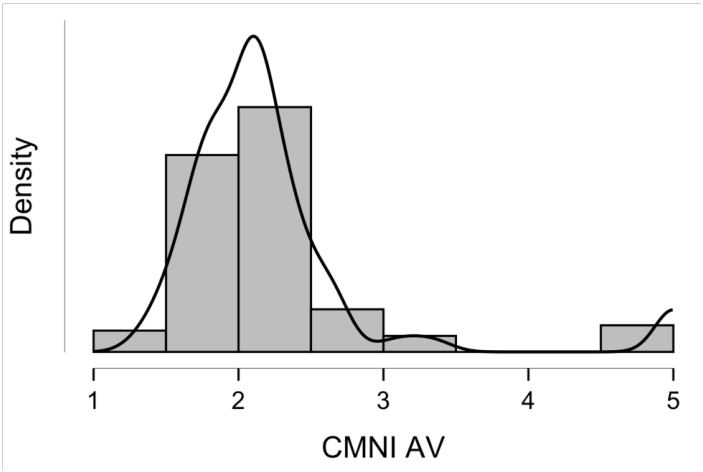


Figure 1: Distribution of individual's average score for Conformity to Masculine Norm (CMNI) a variable

Note. Conformity to Masculine Norm scores could range between 1 and 5, with lower scores indicating lower conformity to masculine norms overall.

[a]CMNI stands for Conformity to Masculine Norms. This is the average across all four of the masculine norms scored in this study.

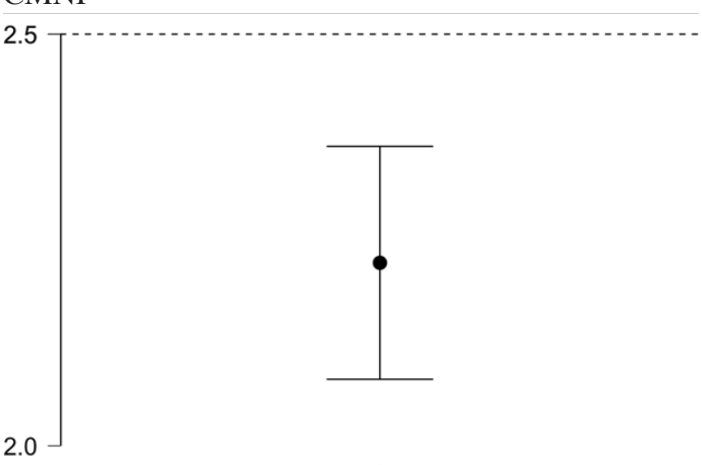


Figure 2: One sample t-test to show significant difference between the sample mean, when compared to the normative sample mean of 2.5 for Conformity to Masculine Norms variable

Note. Conformity to Masculine Norm scores could range between 1 and 5, with lower scores indicating lower conformity to masculine norms overall. 2.5 is therefore considered the normative sample mean.

Attitudes toward sexuality

Descriptive plots displaying density showed a negative skew for attitudes toward sexuality average across the sample (M= 2.367, SD=0.303). Higher ratings on the scale indicate more positive attitudes toward sexuality (See Figure 5).

Likelihood-of-usage for Novel Male Contraceptives

When responses to each individual contraceptive was analyzed, results were varied (See table 2 for general descriptive statistics). RISUG and Nesterone showed non-significant results for likelihood-of-usage when compared to the neutral 3 value, whereby 3 on the Likert scale represented 'neither unlikely nor likely to try'. However, an analysis of likelihood-of-usage for MENT yielded significant results. The one sample t-test for MENT showed a significant likelihood-of-usage (M= [3.505], SD =[1.316]), when compared to the neutral 3 'neither unlikely nor likely to try' value $t(101)=3.855$ and $p<0.001$ with CD = 0.3837 (See Figure 3). The effect size for this analysis was found to be between a small and medium effect size as defined by Cohen's (1988) conventions for a small (d = .20), and medium (d=.50) effect size.

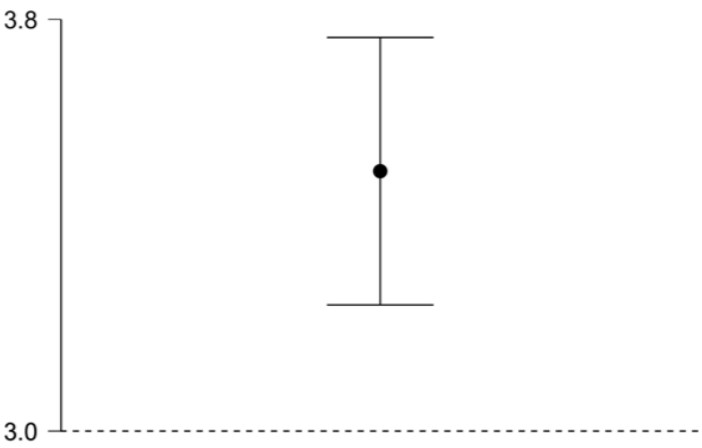


Figure 3: One sample t-test to show significant difference between sample mean and normative sample mean of 3 for likelihood-of-usage of MENT

Note. Likelihood-of-usage scores could range between 1 and 5, with 1 being 'Extremely likely' and 5 being 'Extremely unlikely'. Thus, higher values indicate lower likelihood-of-usage.

Conformity to Masculine norms and Likelihood-of-usage: In order to test the hypothesis that higher conformity to masculine norms will predict less likelihood-of-usage across novel male contraceptives, Pearson correlation tests were conducted. No significant results were seen.

Conformity to Masculine Norms and overall Likelihood-of-usage: There was a non- significant correlation between average Conformity to masculine norms and average likelihood- of-usage across all novel contraceptives $r(101) = 0.075$, $p=0.773$.

Conformity to Masculine Norms and Likelihood-of-usage for RISUG: There was a non- significant correlation between Conformity to masculine norms and likelihood-of-usage the Male Contraceptive RISIG, $r(101) = 0.098$, $p=0.331$.

Conformity to Masculine Norms and Likelihood-of-usage for Nesterone: There was a non- significant correlation between Conformity to masculine norms and likelihood-of-usage the Male

Contraceptive Nesterone, r(101) = 0.055, p=0.585.

Conformity to Masculine Norms and Likelihood-of-usage for MENT: There was a non- significant correlation between Conformity to masculine norms and likelihood-of-usage the Male Contraceptive MENT, r(101) = 0.026, p=0.796.

Self-reliance and its effect on Likelihood of usage of Nestorone

To test the hypothesis that men who endorse 'self-reliance' more highly will favor Nesterone, Pearson correlation tests were conducted. The variables included were Likelihood-of-usage of Nestrone and the self-reliance rating independently of other masculine norm ratings. There was a non-significant correlation between self-reliance and likelihood-of- usage of Nesterone, r(101) = 0.072, p=0.238.

Playboy conformity and its effect on Likelihood-of-usage for MENT

To test the hypothesis that those who endorse the 'Playboy' norm highly will avoid the MENT implant, Pearson correlation tests were conducted. The variables included were Likelihood-of-usage of MENT and the playboy ratings independently of other masculine norm ratings. There was a non-significant correlation between playboy and likelihood-of-usage of MENT, r(101) = -0.089, p=0.189.

Attitudes towards sex and Likelihood-of-Usage

To test the hypothesis that positive attitudes towards sex (higher rating) will predict higher average likelihood of usage of novel male contraceptives, Pearson correlation tests were conducted. The variables included were overall likelihood-of-usage and overall attitudes towards sexuality. There was a non-significant correlation between attitudes towards sexuality and likelihood-of-usage r(89) = -0.0.91, p=0.200.

Comparisons Between Contraceptive Methods

Participants were asked which of the three methods they would most likely use to assess contraceptive preference. Of the sample who answered this question (N=96), Nestrone was chosen most often (47), then RISUG (29) and finally MENT (17). See Figure 4. Descriptive tests were run to assess whether there were differences in likelihood of usage between the three different male contraceptives (Table 2).

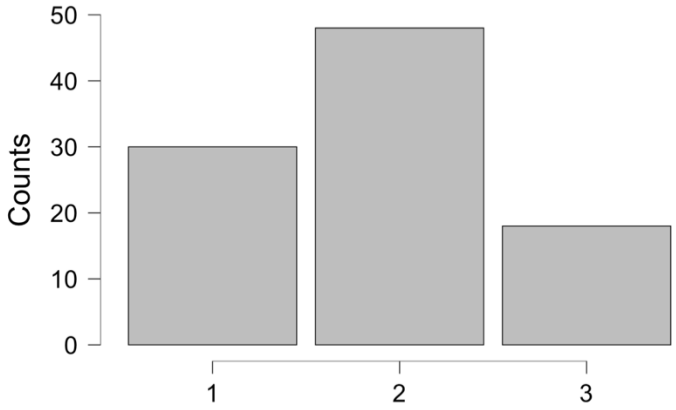


Figure 4: Bar graph showing number individual's preferred method on novel male contraception

Note. Counts indicate number of individuals who chose that contraceptive as their preferred method of novel male contraception.

Reasons for contraceptive preference

The code items for each contraceptive varied depending on the common themes identified by the primary coder. The items and code counts for RISUG can be seen in Table 3, for Nesterone in table 4 and MENT in table 5. It is of note that codes applied three or fewer time could not be included in the Dedoose interrater reliability test, and thus did not yield a Chen's Kappa value (For example, 'No doctors' for Nestorone).

For RISUG, the most common reasoning for preference was "Ease/no hassle/convenience' which accounted for 21.9% of categorized responses (For example, "It seems pretty easy"). 'Longevity' accounted for 19.7% of responses (For example "Last the longest"), 'One-time use' accounted for 18.6% of responses (For example, "I don't have to worry about tit everyday"), and 'Forgetful' accounted for 12.1% of responses (For example, "I would be worried about forgetting to apply the others". The remainder of responses was categorized into 'Reversibility' (9.8%),

'Effective' (8.7%), 'Invasiveness' (5.5%), 'Ejaculate reduction' (2.1%) and 'Minimal side effects' (1%). See Table 6 for details.

For Nestorone the most common reasoning for preference was 'Non-invasive' which accounted for 36.1% of categorized responses (For example, "Lease invasive form since all you're doing is rubbing it on your shoulders"). 'Ease of use' accounted for 21.8% of responses (For example, "Simplest application methods"), 'Non-surgical' accounted for 10.9% of responses (For example, "I would want to avoid implants and surgeries") and 'Most control/flexibility' accounted for 10% (For example, "It has the most day-to-day variability and therefore allows me the most control"). The remainder of responses were categorized into 'Least painful' (6%), 'Easy to stop' (6%), 'Safety' (5%) and 'No Doctors' (2.5%). See table 7 for more detail.

For MENT, the most common reasoning for preference was 'Non-daily use' which accounted for 40% of categorized response (For example, "I don't have to deal with another daily task"). 'Non-Invasive' accounted for 35% of response (For example, "Invasive surgery into scrotum not appealing"), 'Low maintenance accounted for 18.3% (For example "The implant would allow for protection with no maintenance"), and 'Pain free' accounted for the final 6.7% (For example, "Not have to remember to administer something daily and pain free").

Interrater reliability tests all yielded pooled Cohen's kappa values above 0.8, which can be interpreted as very good agreement (Landis & Koch, 1977). Pooled kappa allows for the interpretation of rater agreement across many codes. Inter-rater reliability for RISUG preference reasons had a pooled Cohen's kappa values of .86, with individual code items ranging from 0.78- 0.91(See table 6). Inter-rater reliability for Nesterone preference reasons had a pooled Cohen's kappa value of .87, with individual code items ranging from 0.70-1 (See table 7). Inter-rater reliability for MENT preference reasons had a pooled Cohen's kappa value of .94, with individual code items ranging from 0.83-1. See table 8.

Discussion

General Result Summary

Previous studies have examined more generally how conformity to masculine norms predicts poor health outcomes (Courtenay, 1996; Verbrugge & Wingard, 1987), Specifically, previous literature has focused on how masculine norms discouraging individuals from seeking healthcare professionals and general help (Addis & Mahalik, 2003). This avoidance of good health practices has been found to be further generalizable to an avoidance or misuse of contraceptives (Lohiya et al., 2014). No such studies have included novel male contraceptive preference. Thus, this current study aimed to assess endorsement of specific masculine norms, attitudes towards sexuality, and how both impact preference regarding three real world novel male contraceptives. This body of work contributes to the emerging field of male contraceptive research and identifies reasons behind preferences amongst the population of interest.

The original hypothesis that higher conformity to masculine norms predicts lower likelihood-of-usage for novel male contraceptives was not supported by the data, and the null was accepted. The second hypothesis that positive attitudes towards sex would predict higher average likelihood of usage of novel male contraceptives was not supported by the data, and the null was accepted. Furthermore, both the hypothesis that higher 'Playboy' norm endorsement would predict avoidance of MENT, and higher 'self-reliance' endorsement would predict higher likelihood-of-usage for Nesterone were not supported by the data. The nulls were accepted.

When assessing the primary research question regarding whether conformity to masculine norms could predict likelihood of usage for novel male contraceptives, the results were non-significant. All the correlational analyses run returned non-significant, indicating no relationship between conformity to masculine norms and likelihood-of-usage for novel male contraceptives. What may have contributed to the non-significance of these results is the fact that the sample of participants (predominantly male undergraduate students at a private Southeastern University) did not heavily conform to masculine norms, with a positive skew towards the low endorsement range. The individual's self-reported endorsement of masculine norms was significantly lower than the 'average' value, meaning most individuals had a low

endorsement of masculine norms overall. Therefore, a broad range of masculine norm endorsement was not seen, and a floor effect was noted. A similar skew was seen in the average attitudes towards sexuality scores, with a negative skew towards being observed. A broad range was not observed. Higher ratings on the Sexual Opinion Survey indicate more positive attitudes toward sexuality, and thus the sample displayed predominantly more positive attitudes.

Contraceptive preference

Whilst the original hypotheses were not supported by the data, several discoveries outside of these confines were made. When participants were asked to choose one of the three methods that they would most likely use, a hierarchy of preference was observed. Nestrone was chosen the most often followed by RISUG and finally MENT, with significantly fewer individuals choosing MENT as their preferred method. This indicates a preference for Nestorone, and an avoidance of MENT within this sample.

When assessing likelihood-of-usage for each of the male contraceptives separately, results were mixed. There were no significant results regarding likelihood-of-usage for RISUG and Nesterone; meaning the sample did not have a significant likelihood to use or avoid these contraceptives. However, the sample did show a significant result of their likelihood-of-usage rating regarding MENT, which in fact indicated that this contraceptive was the least appealing to the group. This preference hierarchy could in part be due to the procedures and longevity of each contraceptive. MENT involves a subdermal implant that reduces volume of ejaculate during sexual intercourse. Previous literature has shown that the functionality of the penis is a valued masculine norm, particularly in men who believe that sexual performance is expressed through erections, ejaculation, and penetration (Khann et al., 2008). Thus, this contraceptive impacting ejaculate volume may be seen as a threat to sexual performance. MENT must also be implanted and removed by a healthcare professional. However, it is important to note that this is also the case for RISUG - involving an injection of a non-toxic synthetic chemical into the vas deferens - which did not show the same significant aversions. This suggests that the clinical nature and surgical procedure involved in the MENT implantation did not play as much of a role in the sample's aversion. In terms of longevity, RISUG lasts up to 10 years, MENT lasts one year, and Nestorone involves daily application. It could be suggested that those who would prefer a semi-permanent method that involves no daily upkeep will choose RISUG over MENT because of the frequency MENT would need to be changed; every year as opposed to every decade for RISUG. For individuals who are aversive to semi-permanent methods, Nesterone seems the obvious choice because it involves no clinical invention and can be stopped and started at the user's discretion. In this case, there would be no reason to choose MENT over Nestorone.

Reasonings behind contraceptive preference

The free response answers the question 'Why did you choose this method?' yielded particularly interesting results and common themes for preference were easily identifiable.

Nesterone: The most common reason given for the preference of Nesterone was that it was 'non-invasive' (For example "Least invasive form since all you're doing is rubbing it on your shoulders''). As previously mentioned, Nestorone is the only contraceptive of the three that does not involve a procedure of any kind. 'Non-invasiveness' seems particularly salient reason for preference; this finding stands to reason given the target population given the generalized avoidance of healthcare professionals amongst men (Addis & Mahalik, 2003). Thus, providing a non-invasive method aligns with the values of the target population. The second most common reason for Nesterone preference was 'Ease of use', with many citing the ease of application (For example, "Simplest application method"). Consequently, it seems important to ensure that the contraceptive application is simple as possible. Furthermore, 10% of individuals in this category cited "Most control" as a reason for their preference (For example, "It has the most day-to-day variability and therefore allows me the most control"). With Nesterone being the only self-applied contraceptive, this seems to be the method of choice for those who prefer to have daily autonomy over their reproductive capabilities or who don't wish to commit to a 1-year or 10-year contraceptive. This is further

supported by the 6% of responses that cited Nestorone as being 'Easy to stop'. Hence, the option for quick and easy reversibility is another crucial consideration for novel contraceptive design.

RISUG: For those who preferred RISUG, the most common reasoning was found to be "Ease/No hassle/Convenience', followed by 'Longevity'. RISUG is the longest lasting of the three methods - with an effective period of 10 years - and thus for those looking to avoid pregnancy without the daily application of Nestorone, or yearly change for MENT, RISUG seemed to be the most suitable option. Furthermore, these individuals were aware of the invasive procedure involved in RISUG application, however this was outweighed by 'Reversibility', which accounted for almost 10% of the reasons for RISUG preference. It is also of note that many individuals indicated that their preference of RISUG over Nesterone was down to daily-use factor; upwards of 12% of the individuals cited that they would forget to apply a daily gel (For example, "I would be worried about forgetting to apply the others"). With RISUG being the second most preferred method of the three, it seems to suit individuals who do not mind the invasiveness of the original procedure and do not want the daily hassle of remembering their contraceptive.

MENT: For those who preferred MENT, the most common reasoning was found to be a combination of 'Non-daily use' (For example, "I don't have to deal with another daily task") and 'Less invasive' (For example, "Invasive surgery into scrotum not appealing"). The individuals who preferred MENT seem to fit a niche between RISUG and Nesterone; whilst they do not wish to partake in the daily application of Nestorone, they are also opposed to the procedure involved with RISUG. Although MENT also involves a surgical procedure, it seems that the upper arm implant is considered less invasive to these individuals than an injection into the vas-deferens of the penis. Previous literature that has shown functionality of the penis to be valued masculine norm particularly in men who believe that sexual performance is expressed through erections, penetration (Khann et al., 2008). Whilst RISUG has no known negative side effects in regard to the functionality of the penis or sexual performance, the mere notion of interference with the penis may be seen as a threat to this valued masculine norm.

Despite the null hypothesis having been accepted, these findings regarding male contraceptive preference are of significant importance regarding future research efforts. The above data has shown that the sample showed preference concerning the novel forms of male contraceptives presented. The reasons indicated by the sample regarding their preference indicate specific driving factors driven by what those individuals' value. Namely ease of use, longevity and surgical procedure avoidance. Thus, when funding research into new male contraceptive measures, understanding these valued preferences could ensure the largest uptake and market possible.

Strengths

This study examined a reasonable sample of the target population, with 103 participants completing the entire survey. With the sample pool being predominantly Duke affiliated students, this could enable results to be generalized to other small, private, Universities in the American South. The electronic survey disseminated by the researcher was anonymous, and no identifiable data was collected. Anonymity and consequent confidentiality of responses minimizes the influence of demand characteristics and allowed participants to answer as truly as they wished. Furthermore, this study utilized two validated measures to operationalize the independent variables: namely the CMNI-46 and The Sexual Opinion Survey. These are both validated and reliable surveys that have been used in many previous studies. Both sets of questions had an internal consistency above 0.70, meaning that each measure accurately operationalized the variable in question and ensured high sensitivity of each individual item.

Limitations

This survey utilized in this study was disseminated by the undergraduate researcher to predominantly male undergraduate students, and thus the WEIRD paradigm cannot be ignored. University students tend to be white, educated, industrialized, rich and democratic and thus the data is non-representative of the global mean. The homogeneity of the sample was seen during data analysis; the individuals self-reported endorsement of masculine norms was significantly

lower than the 'average' value, meaning most individuals had a low endorsement of masculine norms overall. The data point for conformity to masculine norms was clumped towards the low endorsement range, which may in part explain the non-significant results when relating masculine endorsement to contraceptive usage. In order to get a broader range of masculine endorsement values, a new inventory or scale should be considered. Whilst the CMNI-46 is a well validated scale, it was created in 2009 over 12 years ago meaning that it is possible the measures are outdated for modern day society. Since the sample consisted of predominantly college students at a liberal-arts university, they may well reject stereotypical forms of masculinity and thus the questions relating to masculinity were too overt to yield a broad range of answers. The CMNI-46 should either be validated with this population of interest, or a new inventory reflecting modern day notions of masculinity should be created and validated.

Furthermore, this study saw an attrition rate of 39.1% for individuals who were not screened out of the survey but did not finish the survey to completion. This could be in part due to the nature of the survey questions, including masculinity, sexuality and contraception; all of which can be sensitive topics for individuals. Many of those who did not complete the survey ended their survey at the very first question which involved a description of the contraceptive RISUG, which included terms such as 'semen' and 'scrotum', which may have been potentially off putting to some. Others who ended the survey early tended to do so before answering the first question assessing attitudes towards sexuality ("If i could I would frequently change sexual partners"). Given the sensitive and often taboo nature of sex and sexuality, this may be a result of stress associated with talking about topics pertaining directly to sex. Overall attrition could in part be due to survey fatigue. All survey responses were voluntary, and with a total of 49 questions, participants may not have felt compelled to finish the survey to completion. In order to address this, some form of compensation for those who complete the survey may be necessary to encourage completion. Researchers may also wish to communicate how the data will be used and securely stored as a means of being transparent and putting participants at ease given the nature of the questions.

Implications

Contraception is the most effective way to avoid unplanned pregnancies (Sonfield et al., 2013) – and while women still account for 70% of all contraceptive use, there remain only two forms of contraception for men. If men and women were able to share the reproductive burden of family planning, male contraceptives would have a measurable impact on the unplanned pregnancy rate (Dorman, 2018). With RISUG MENT and Nesterone all in varying stages of clinical trials, identifying ways in which to integrate these methods into the reproductive landscape is paramount to maximizing uptake and usage. The ultimate goal of this research study was to better understand preference regarding these novel male contraceptives. By identifying a hierarchy of preference between these three methods, and additional qualitative reasonings behind individual choice, this study yields important data that help inform future contraceptive design and research to meet the needs and preferences of the demographic. Furthermore, with all three of the contraceptives being chosen as the preferred method by numerous individuals, this indicates that one contractive will not fit all. Future research will need to encompass multiple methods of male contraceptives to satisfy a broader range of potential users. In turn, this may increase male contraceptive use, and have profound public health and financial impacts given the potential to prevent unwanted pregnancies.

Future Studies

As previously mentioned, the sample utilized for this study was limited by its homogenous nature, predominantly undergraduate students. In future studies, a larger more diverse demographic would be preferable to ensure a better understanding of cross-cultural preferences. This could include men from other countries of reproductive age, such as men from British Universities. This would aid a wider perspective on male contraceptive preference than afforded by this limited sample. Additionally, whilst the CMNI-46 is a well validated scale, it did not yield a broad range of values for the target population. Further validation of this measure is needed for present day individuals of reproductive age; the original survey was created in

2009, which whilst only 12 years ago may not be valid across younger male age groups. Furthermore, this survey's methodology was predominantly quantitative and correlational, meaning relationships between variables should not be interpreted causally. Whilst one qualitative measure was included, this yielded the most interesting and applicable findings. Thus, utilizing interviews and focus groups in the future would help provide further insight into the nuances involved in men's contraceptive decision-making. This would enrich the data yield and more usefully guide specific policy changes and research directions regarding male contraception.

References

1. Addis, M. E., & Mahalik, J. R. (2003). Men, masculinity, and the contexts of help seeking. American Psychologist, 58(1), 5–14. doi: 10.1037/0003-066x.58.1.5
2. Anawalt, B. D., Roth, M. Y., Ceponis, J., Surampudi, V., Amory, J. K., Swerdloff, R. S., ... & Wang, C. (2019). Combined nestorone–testosterone gel suppresses serum gonadotropins to concentrations associated with effective hormonal contraception in men. Andrology, 7(6), 878-887
3. Anderson, R. A., & Baird, D. T. (2002). Male contraception. Endocrine reviews, 23(6), 735-762.
4. Banks, I. (2004). New models for providing men with health care
5. Bearak, J., Popinchalk, A., Alkema, L., & Sedgh, G. (2018). Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a Bayesian hierarchical model. The Lancet Global Health, 6(4), e380-e389.
6. Bem, S. L. (1981). The BSRI and gender schema theory: A reply to Spence and Helmreich.
7. Bothmer, M. I. K. V., & Fridlund, B. (2005). Gender differences in health habits and in motivation for a healthy lifestyle among Swedish university students. Nursing and Health Sciences, 7(2), 107–118. doi: 10.1111/j.1442-2018.2005.00227.x
8. Chao, J., Page, S. T., & Anderson, R. A. (2014). Male contraception. Best Practice & Research Clinical Obstetrics & Gynaecology, 28(6), 845-857.
9. Courtenay, W. H. (2000). "Constructions of Masculinity and Their Influence on Men's Well-Being: A Theory of Gender and Health." Social Science & Medicine 50 (10): 1385–1401. doi:10.1016/ S0277-9536(99)00390-1
10. Courtenay W. H. (2000). Constructions of masculinity and their influence on men's well-being: a theory of gender and health. Social science & medicine (1982), 50(10), 1385–1401. https://doi.org/10.1016/s0277-9536(99)00390-1
11. Courtenay, W. H. (2000). Engendering health: A social constructionist examination of mens health beliefs and behaviors. Psychology of Men & Masculinity, 1(1), 4–15. doi:10.1037/1524-9220.1.1.4
12. Dorman, E., & Bishai, D. (2012). Demand for male contraception. Expert review of pharmacoeconomics & outcomes research, 12(5), 605-613
13. Dorman, Et al. "Modeling the Impact of Novel Male Contraceptive Methods on Reductions in Unintended Pregnancies in Nigeria, South Africa, and the United States." Contraception 2018
14. Eberhardt, J., Van Wersch, A., & Meikle, N. (2009). Attitudes towards the male contraceptive pill in men and women in casual and stable sexual relationships. BMJ Sexual & Reproductive Health, 35(3), 161-165.
15. Fallis, E., Gordon, C., & Purdon, C. (2011). Sexual anxiety scale. Handbook of sexuality-related measures, 228-231.
16. Fisher, W. A. (1984). Predicting Contraceptive Behavior Among University Men: The Role of Emotions and Behavioral Intentions 1. Journal of Applied Social Psychology, 14(2), 104-123.
17. Fisher WA, Byrne D, White LA, Kelley K. Erotophobia-erotophilia as a dimension of personality. The Journal of Sex Research. 1988;25:123–151.
18. Fisher WA, Grenier G, Watters WW, Lamont J, Cohen M, Askwith J. Students' sexual knowledge, attitudes toward sex, and willingness to treat sexual concerns. Journal ofMedical Education. 1988;63:379–385.
19. Garfield, C. F., Isacco, A., & Rogers, T. E. (2008). A review of men's health and masculinity. American Journal of Lifestyle Medicine, 2(6), 474-487.
20. Ilani, N., Roth, M. Y., Amory, J. K., Swerdloff, R. S.,

Dart, C., Page, S. T., Bremner, W. J., Sitruk-Ware, R., Kumar, N., Blithe, D. L., & Wang, C. (2012). A new combination of testosterone and nestorone transdermal gels for male hormonal contraception. The Journal of clinical endocrinology and metabolism, 97(10), 3476–3486. https://doi.org/10.1210/jc.2012-1384

21. Iwamoto, D. K., Cheng, A., Lee, C. S., Takamatsu, S., & Gordon, D. (2011). "Man-ing" up and getting drunk: The role of masculine norms, alcohol intoxication and alcohol-related problems among college men. Addictive behaviors, 36(9), 906-911.
22. Kumar, N., Fagart, J., Liere, P., Mitchell, S. J., Knibb, A. R., Petit-Topin, I., Rame, M., El-Etr, M., Schumacher, M., Lambert, J. J., Rafestin-Oblin, M. E., & Sitruk-Ware, R. (2017). Nestorone® as a Novel Progestin for Nonoral Contraception: Structure-Activity Relationships and Brain Metabolism Studies. Endocrinology, 158(1), 170–182. https://doi.org/10.1210/en.2016-1426
23. Khan, S. I., Hudson-Rodd, N., Saggers, S., Bhuiyan, M. I., Bhuiya, A., Karim, S. A., & Rauyajin, O. (2008). Phallus, performance and power: crisis of masculinity. Sexual and Relationship Therapy, 23(1), 37-49.
24. Khilwani, B., Badar, A., Ansari, A. S., & Lohiya, N. K. (2020). RISUG® as a male contraceptive: journey from bench to bedside. Basic and clinical andrology, 30, 2. https://doi.org/10.1186/s12610-020-0099-1
25. Lacasse, K., & Jackson, T. E. (2019). Conformity to masculine norms predicts US men's decision-making regarding a new male contraceptive. Culture, Health & Sexuality, 1–17. doi: 10.1080/13691058.2019.1658806
26. Limmer, M. 2016. "'I Don't Shag Dirty Girls': Marginalized Masculinities and the Use of Partner
27. Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics, 363-374.
28. Lipsey, Richard G., Kenneth Carlaw, and Clifford Bekar. Economic Transformations: General Purpose Technologies and Long-Term Economic Growth. Oxford; New York: Oxford University Press
29. Lohiya, N. K., Alam, I., Hussain, M., Khan, S. R., & Ansari, A. S. (2014). RISUG: an intravasal injectable male contraceptive. The Indian journal of medical research, 140(Suppl 1), S63.
30. Luyt, R. (2018). Masculinities Representations Inventory (MRI, English Version): A measure of gender (re) presentation. The Journal of Men's Studies, 26(2), 157-183.
31. Macapagal, K. R., & Janssen, E. (2011). The valence of sex: Automatic affective associations in erotophilia and erotophobia. Personality and individual differences, 51(6), 699-703.
32. Mahalik, J. R., Locke, B. D., Ludlow, L. H., Diemer, M. A., Scott, R. P., Gottfried, M., & Freitas, G. (2003). Development of the conformity to masculine norms inventory. Psychology of Men & Masculinity, 4(1), 3.
33. Martínez, I. V., & Ramos, A. (2018). The perspectives of men in Mexico City about vasectomy
34. Mechanic, D., & Cleary, P. D. (2004, February 9). Factors associated with the maintenance of positive Health behavior. Retrieved from https://www.sciencedirect.com/science/article/pii/0091743580900237?via=ihub
35. Parent, M. C., and B. Moradi. 2009. "Confirmatory Factor Analysis of the Conformity to Masculine Norms Inventory and Development of the Conformity to Masculine Norms Inventory-46." Psychology of Men & Masculinity 10 (3): 175–189. doi:10.1037/a0015481
36. Qureshi, B. (2001). Book reviews : Improving men's health By Colin Francome. Published by Middlesex University Press, London 2000. Paperback 262 pp. Price £14.99 ISBN 1898-253-36-6. Journal of the Royal Society for the Promotion of Health, 121(1), 65–65. https://doi.org/10.1177/146642400112100118
37. Rye, B. J., Meaney, G. J., & Fisher, W. A. (2011). Sexual opinion survey. Handbook of sexuality-related measures, 231-236.
38. Selection as a Sexual Health Risk Reduction Strategy in Heterosexual Young Men." American Journal of Men's Health 10 (2): 128–140.
39. Snell WE Jr, Papini DR. The Sexuality Scale: An instrument to measure sexual-esteem, sexual-depression, and sexual-preoccupation. J Sex Res. 1989; 26:256–63

40. Sonfield, A., Hasstedt, K., Kavanaugh, M. L., & Anderson, R. (2013). The social and economic benefits of women's ability to determine whether and when to have children.
41. Sundaram, K., Kumar, N., & Bardin, C. W. (1993). 7α-Methyl-nortestosterone (MENT): the optimal androgen for male contraception. Annals of medicine, 25(2), 199-205.
42. Thom, B. (2003). Risk-taking behaviour in men: Substance use and gender. HDA.
43. Tudiver, F., & Talbot, Y. (1999). Why don't men seek help? Family physicians' perspectives on help- seeking behavior in men. Journal of Family practice, 48, 47-52.
44. Verbrugge, L. M., Wingard, D. L., & Features Submission, H. C. (1987). Sex differentials in health and mortality. Women & health, 12(2), 103-145.
45. Von Eckardstein, S., Noe, G., Brache, V., Nieschlag, E., Croxatto, H., Alvarez, F., ... & Sundaram, K. (2003). A clinical trial of 7α-methyl-19-nortestosterone implants for possible use as a long-acting contraceptive for men. The Journal of Clinical Endocrinology & Metabolism, 88(11), 5232- 5239.
46. Waldron, I. (1988). Gender and health-related behavior. In Health behavior (pp. 193-208). Springer, Boston, MA
47. Walker, S. (2011). Attitudes to a male contraceptive pill in a group of contraceptive users in the UK. Journal of Men's Health, 8(4), 267-273.
48. Winter L. (1988). The role of sexual self-concept in the use of contraceptives. Family planning perspectives, 20(3), 123–127.
49. Yarber, W. L., Milhausen, R. R., Crosby, R. A., & Torabi, M. R. (2005). Public Opinion About Condoms for HIV and STD Prevention: A Midwestern State Telephone Survey. Perspectives on Sexual and Reproductive Health, 37(3), 148–154. doi: 10.1363/3714805

Tables

Table 1a–*Summary statistics table for each of the four masculine norms and over all conformity to masculine norm average*

	Number of complete responses	Mean Rating	Std. Deviation	Minimum Rating	Maximum Rating
Playboy	103	2.437	0.874	1	5
Self-Reliance	103	2.375	0.763	1	5
Heterosexual presentation	103	2.223	0.887	1	5
Power-Over-Women	98	1.653	0.530	1	3.250
CMNI[a] average	103	2.222	0.723	1	5

Note. Number of completed responses indicates number of individuals who completed all of the questions related to each specific norm. Scores could range between 1 and 5, with 5 being highest conformity value.

[a]CMNI stands for Conformity to Masculine Norms. This is the average across all four of the norms.

Table 1b–*Summary of statistics for the Sexual Opinion Survey*

	Number of complete responses	Mean Rating	Std. Deviation	Minimum Rating	Maximum Rating
Sexual Opinion Survey	89	2.367	0.303	1	2.905

Note. Scores could range between 1 and 4, with higher ratings indicating more positive attitudes toward sexuality.

Table 1c–*Summary of statistics for 'Likelihood-of-usage' self-ratings for all three contraceptives*

	Number of complete responses	Mean Rating	Std. Deviation	Minimum Rating	Maximum Rating
Likelihood-Of-Usage	101	3.089	1.109	1	5

Table 2–*Likelihood-of -usage ratings for each of the three contraceptives*

	Likelihood-of-usage		
	RISIG	**Nesterone**	**MENT**
Number of individuals who chose this as their preferred method	30	48	18
Mean Likelihood-of-usage rating	2.989	3.007	3.148
Std. Deviation	1.105	1.075	1.15
Minimum rating	1	1	1.333
Maximum rating	5	5	5

Note. Likelihood-of-usage ratings could range from 1 to 5, with 1 being 'Extremely likely' and 5 being 'Extremely unlikely'

Table 3–*Coded items for RISUG preference reasoning*

Media \ Codes	Ease/no hassle/Convenience	Effective	Ejaculate reduction	Forgetful	Invasive	Longevity	One time use	Reversibility	minimal side effects	Totals
RISIG words copied and pasted	20	8	2	11	5	18	17	9	1	91
Totals	20	8	2	11	5	18	17	9	1	

Table 4–*Coded items for Nesterone preference reasoning*

Media \ Codes	Ease of use	Easy to stop	Least painful	Most control/flexibility	No Doctors	Non-invasive	Non-surgical	Safety	Totals
Nestrone text copied and pasted	26	8	8	12	3	43	13	6	119
Totals	26	8	8	12	3	43	13	6	

Table 5–*Coded items for MENT preference reasoning*

Media \ Codes	Less invasive	Low maintenance	Non-daily use	Pain-free	Totals
MENT copied and pasted	21	11	24	4	60
Totals	21	11	24	4	

Table 6–*Summary of Interrater agreement and number of items per RISUG preference facet*

RISUG item	Cohen's Kappa	Item %	Example
Ease/no hassle/convenience	0.87	21.9	It seems pretty easy
Longevity	0.86	19.7	Lasts the longest
One-time use	0.78	18.6	I don't have to worry about it every day
Forgetful	0.1	12.1	I would be worried about forgetting to apply the others
Reversibility	0.88	9.8	Easily reversible
Effective	0.88	8.7	100% effective
Invasive	0.79	5.5	My biggest worry with the RISUG though is the invasiveness.
Ejaculate Reduction	--	2.1	Reduction in ejaculate
Minimal side effects	---	1	Seems like it has the least side effects

Table 7–*Summary of Interrater agreement and number of items per Nesterone preference facet*

Nesterone item	Cohen's Kappa	Item %	Example
Non-invasive	0.89	36.1	Least invasive form since all you're doing is rubbing it on your shoulders
Ease of use	0.9	21.8	Simplest application method
Non-surgical	0.77	10.9	I would want to avoid implants and surgeries.
Most control/ flexibility	0.7	10	It has the most day-to-day variability and therefore allows me the most control.
Easy to stop	1	6	It seems as if it would be the easiest to stop.
Least painful	1	6	I'd rather choose the least painful method
Safety	1	5	Seems like the most safe option
No doctors	---	2.5	Doesn't require a trip to the doctor

Table 8–*Summary of Interrater agreement and number of items per MENT preference facet*

MENT items	Cohen's Kappa	Item %	Example
Non-daily use	0.91	40	I don't have to deal with another daily task
Less invasive	1	35	Invasive surgery into scrotum not appealing
Low maintenance	0.83	18.3	The implant would allow for protection with no maintenance.
Pain-free	1	6.7	Not have to remember to administer something daily and pain free

Appendix

Demographic screening questions:

1) "Do you ejaculate (sperm, semen, spermatozoid)?" – Y or N. Yes, Include. If No, exclude
2) Do you have sex with a partner who can get pregnant either exclusively or otherwise – Y or N
3) Are you sexually active – Y or N.
4) What is your gender? Male, Transgender Female, Transfeminine, Assigned Male At Birth (AMAB), a Person with Disorder of Sex Development (intersex), Option to write in _____.

DV–The items below inquire about some of your attitudes and opinions towards Real world Male Contraceptives. You will be given a brief paragraph explaining each contraceptive method and then asked some follow up questions.

1. **RISUG**–works by an injection of a non-toxic synthetic chemical into the vas deferens; local anesthetic is administered; an incision is made in the scrotum.
   The chemical is effective almost immediately after it is injected. The chemical stays in place until a man decides that he wants to have children. It can then be washed out using another injection which dissolves and flushes it out.
   It is 99.9% effective, reversible, cheap, can last 10 years if wanted. No reduction in ejaculation, no widespread side effects however does not protect against STDs.

   How likely would you be to use the male contraceptive RISUG?
   1–Very Unlikely 2–Unlikely 3–Neutral 4–Likely 5–Very Likely

2. **Nestrorone**–a contraceptive gel that is applied once daily and absorbed through the skin of a man's shoulders. 99.9 % effective at prevents pregnancy. Early research findings show that the Nestorone®/ Testosterone gel is well tolerated with no serious adverse events. Effects can be immediately reversed if gel is not rubbed on the shoulders, with no effect on fertility however does not protect against STDs.

   How likely would you be to use the male contraceptive Nestorone®?
   1–Very Unlikely 2–Unlikely 3–Neutral 4–Likely 5–Very Likely

3. **MENT**–a one-year implant that is placed under the skin of the upper arm. MENT is effective, safe, acceptable, long acting, and reversible requires an approach that will reduce sperm count and ejaculate volume without requiring frequent administration (e.g., daily use). Can be removed however does not protect against STDs.

   How likely would you be to use the male contraceptive MENT?
   1–Very Unlikely 2–Unlikely 3–Neutral 4–Likely 5–Very Likely

Comparison –
Of the three above mentioned male contraceptives, which would you be most likely to use/choose?
1. RISUG (injection of a non-toxic synthetic chemical into the vas deferens)
2. Nesterone (a contraceptive gel that is applied once daily to shoulder)
3. MENT (one-year implant)

Why did you choose this method?

CMNI adaptation

Thinking about your own actions, feelings and beliefs, please indicate how much you personally agree or disagree with each statement by selecting Strongly Disagree, Disagree, Agree, or Strongly agree. There are no right or wrong responses to the statements. You should give the responses that most accurately describe your personal actions, feelings and beliefs. It is best if you respond with your first impression when answering.

0 (strongly disagree) to 4 (strongly agree)

1. If I could, I would frequently change sexual partners
2. I hate asking for help
3. Being thought of as gay is not a bad thing
4. I ask for help when I need it
5. I would only have sex if I was in a committed relationship
6. I would be furious if someone thought I was gay
7. It would not bother me at all if someone thought I was gay
8. In general, I control the women in my life
9. I would feel good if I had many sexual partners
10. It would be awful if people thought I was gay
11. I never ask for help
12. Women should be subservient to men
13. It would be enjoyable to date more than one person at a time
14. I would feel uncomfortable if someone thought I was gay
15. I am not ashamed to ask for help
16. Things tend to be better when men are in charge
17. It bothers me when I have to ask for help
18. I love it when men are in charge of women
19. I try to avoid being perceived as gay

Sexual opinion survey

Thinking about your own actions, feelings and beliefs, please indicate how much you personally agree or disagree with each statement by selecting Strongly Disagree, Disagree, Agree, or Strongly agree Please respond to each item as honestly as you can. There are no right or wrong answers on your answers will be completely anonymous.

1. I think it would be very entertaining to look at hard-core erotica.
2. Erotica is obviously filthy, and people should not try to describe it as anything else.
3. Swimming in the nude with the opposite sex would be an exciting experience.
4. Masturbation can be an exciting experience.
5. If I found that a close friend of mine was homosexual, it would annoy me.
6. If people thought I was interested in the opposite sex, I would be embarrassed.
7. Engaging in group sex is an entertaining idea.
8. I personally find that thinking about engaging in sexual intercourse is arousing.
9. Seeing an erotic movie would be sexually arousing to me.
10. Thoughts that may have homosexual tendencies would not worry me.
11. The idea of being physically attracted to members of the same sex is not depressing.
12. Almost all sexually explicit material is nauseating.

1. It would be emotionally upsetting to me to see someone exposing themselves publicly.
2. Watching an erotic dancer of the opposite sex would not be very exciting.
3. I would not enjoy seeing an erotic movie.
4. When I think about seeing pictures showing someone of the same sex as myself masturbating, it nauseates me.
5. The thought of engaging in unusual sexual practices is highly arousing.
6. Manipulating my genitals would probably be an arousing experience.
7. I do not enjoy daydreaming about sexual matters.
8. I'm not curious about explicit erotica.
9. The thought of having a long-term sexual relationship with more than one partner is not disgusting to me.

Utilizing Patient Characteristics to Predict the Likelihood of Heart Condition

Jackson Cathey, Carson Herman, Michelle Tetro, Melody Gao

Volume 1, Issue 1

Utilizing Patient Characteristics to Predict the Likelihood of Heart Condition

With focus on sex, resting ECG results, and exercise-induced angina

Jackson Cathey, Carson Herman, Michelle Tetro, Melody Gao

Department of Statistics, Duke University
https://doi.org/10.55894/dv1.12

Abstract

As primary care providers deal with a multitude of chronic conditions, cardiovascular disease may often be overlooked. Finding cardiovascular disease early in a patient increases treatment options and reduces the likelihood of further complications. Thus, it is important that providers have tools at their disposal to assess patients' odds of heart disease, given past medical histories and vital signs in clinic. Utilizing patient data from the Cleveland Heart Clinic, we developed a predictive model for assessing the odds that a patient evaluated in clinic may have an underlying heart condition. We selected each variable based on its relative capacity for prediction of heart condition, then further analyzed the contributes of this feature in the model. The final model includes the sex of the patient, their self-reported experience of exercise-induced angina, and resting ECG results. This model can be used in the clinic to estimate a patient's odds of heart disease. Unlike current models used in primary care clinics, this model uses fewer variables and specifically assesses the odds of a current condition, as opposed to the future development of cardiovascular disease. Testing directly for heart disease is costly and time-consuming, so it's useful to build a predictive diagnostic model.

Introduction

Coronary heart disease develops when coronary arteries are narrowed, unable to transport the necessary oxygenated blood throughout the body (Libby et al., 2005). The high prevalence of the chronic cardiovascular disease calls for early detection. Accounting for one of every four death, cardiovascular disease stands as the leading cause of death in the United States (Murphy et al., 2018). According to the CDC, 47% of Americans meet at least one of three key risk factors for heart disease: high blood pressure, high blood cholesterol, and smoking (CDC, 2021). As medicine progresses to capitalize on preventative measures, it is imperative that heart disease is predicted early on in order to avoid the development of such complications. Currently, there is a range of coronary heart disease risk factors that are analyzed to arrive at a diagnosis: hypertension, high levels of low-density lipoprotein cholesterol, low levels of high-density lipoprotein cholesterol, high total cholesterol, high triglycerides, diabetes mellitus, smoking, obesity, aging, gender, physical inactivity, age, socioeconomic and psychological stress, family history, and more (Marateb et al. 2015). Diagnosis can be both difficult and complex. While the modern technology developed aids the process, there are still oversights in diagnosing cardiovascular disease. However, if healthcare physicians can utilize universal, concrete variables to predict which patients may be at a greater risk for cardiovascular disease, then this leading cause of morbidity and mortality can be caught early in patients and prevented. Utilizing data collected by the Cleveland Clinic Foundation, variables of heart disease

that are analyzed to arrive at a diagnosis: hypertension, high levels of low-density lipoprotein cholesterol, low levels of high-density lipoprotein cholesterol, high total cholesterol, high triglycerides, diabetes mellitus, smoking, obesity, aging, gender, physical inactivity, age, socioeconomic and psychological stress, family history, and more (Marateb et al. 2015). Diagnosis can be both difficult and complex. While the modern technology developed aids the process, there are still oversights in diagnosing cardiovascular disease. However, if healthcare physicians can utilize universal, concrete variables to predict which patients may be at a greater risk for cardiovascular disease, then this leading cause of morbidity and mortality can be caught early in patients and prevented. Utilizing data collected by the Cleveland Clinic Foundation, variables of heart disease will be compared to condition outcome, developing an preliminary predictive model for a given patient's risk of heart condition.

Data and Collection

A group of colleagues who work closely with cardiovascular disease (Andras Jonosi, M.D., William Steinbrunn, Matthias Pfisterer, M.D., and Robert Detrano, M.D., Ph.D) gathered clinical data that has been compiled into one data set made available to the public (Jonosi et al, 1984). These researchers assessed quantitative and qualitative variables of patient groups undergoing angiography at a range of hospitals in the United States. Patients with prior infarction, valvular disease, and prior catheterization were excluded from the data collection. The original data set had 76 features. However, 14 clinical variables were chosen as key variables to analyze possible risk factors for cardiovascular disease after a brief statistical analysis. The researchers gathered several attributed in addition to 3 noninvasive tests as a part of the research protocol, including exercise electrocardiogram, thallium scintigraphy, and cardiac fluoroscopy. Additionally, the invasive gold standard test, coronary angiogram, was conducted. The data set can be used for testing proposed systems for cardiovascular disease. We filtered the data in this study to the include 297 patients from 303 original records of the Cleveland Clinic who received angiography between May 1981 and September 1984. Each patient ID includes the individual patients' vitals

The 14 clinical variables were chosen by Dr. Jonosi based on their evident links to cardiovascular disease. The first variable recorded was the patient's age in years. Age can be classified as an important risk factor with heart disease because studies show that the risk begins early in adult life but the probability of developing the disease increases as individuals get older (Franklin et al., 2001). The sex of the patient is another relevant variable regarding cardiovascular disease, as recent literature has shown that males might be at higher risk as a population. Furthermore, studies have revealed that there are subtle differences in the presentation of symptoms of coronary heart disease for males and females, with males exhibiting increased complications (Milner et al., 1999). It was also assessed whether patients reported endorsing chest pain as a variable, as chest pain is a patient-reported symptom that may reflect when the heart is under distress, ranging from typical angina to atypical angina to non-anginal pain to asymptomatic symptoms. Chest pain can either result from a cardiac origin or noncardiac causes, such as psychological disorders (Lenfant, 2010). Despite this ambiguity, it is a well-document predictor of cardiovascular disease that can be collected without testing, rendering it an effective variable to be used in primary care clinics across the nation. The resting blood pressure in mm Hg on admission to the hospital is another variable gathered in the data. Since high blood pressure can be attributed to cardiovascular disease, hypertension is a likely precursor to coronary artery disease and associated heart conditions. Beyond these simple vitals and patient descriptors laboratory blood work can offer great insight to the biochemical underpinnings of coronary artery disease. Serum cholesterol levels—measured in mg/dL—reflect how dietary intake of high-density lipoprotein cholesterol can impact the heart, as it can accumulate in the arteries over time (Blesso et al., 2018). In addition, fasting blood sugar can be an indication of insulin levels. Diminished endogenous insulin or development of insulin resistance can place stress on the heart.

The variables gathered from the noninvasive tests analyze heart function, extending beyond simple patient intake data. Electrocardiographic results are collected by assessing the electrical signals of one's heart. ECG's can be conducted in clinic by placing electrodes on patients' chests, yielding readings of

heart function. A physician analyzing ECG results can determine the length it takes for the electrical wave to pass through the heart, classifying such activity as normal or irregular. Studies show that baseline ECG abnormalities can be associated with cardiovascular disease mortality (Bacquer et al., 1998). Collecting data on the maximum heart rate achieved (like high blood pressure) is integral in understanding the stress placed on the heart in distributing oxygenated blood. Elevated resting heart rate indicates possible problems, as the heart may be working harder than expected for a healthy individual. Elevated heart rate thus has been shown to be associated with cardiovascular disease and higher risk of mortality (Cooney et al., 2009). Exercise-induced angina is operationalized as chest pain (of ranging severity) experienced when patients are undergo exercise testing. One study found that there is a correlation between abnormal increase in pulmonary artery pressure and slight reduction in stroke volume when patients with heart disease exercise (Wiener et al. 1968). Therefore, exercise-induced angina has been reported as a physical indicator of underlying heart disease. Researchers also collected ST depression (the electrically neutral area of the complex between ventricular depolarization and repolarization) induced by exercise relative to rest during treadmill ECG stress tests. Cardiac fluoroscopy offers insight into the vessels impacted by heart disease. The number of vessels colored can indicate vessels afflicted. Through thallium scintigraphy, the amount of blood reaching each ventricle of a patient's hearts can be measured. The thallium stress test includes an exercise and resting portion to measure the blood flow through different stages, indicating heart performance in a range of environments.

The final variable collected in this data matrix is the condition of the patient: whether they were diagnosed with coronary heart disease or not. This variable is essential for comparing the patients' results with the other features in the data set, as it places indicators of cardiovascular disease within the context of one's diagnosis. The condition of the patient is essential for moving forward with the research question of analyzing—based on this dataset—which concrete variables are most strongly associated a coronary heart disease diagnosis.

Methodology

The 14 different variables collected in the Heart Disease Cleveland UCI data set were analyzed. The data set includes 297 patients, 201 men and 96 women, the mean age is 56 years old (sd = 9.1), and 137 (46%) patients with diagnosed heart disease. The mode exercise-induced chest pain type is group 2, non-anginal pain. The average resting blood pressure was 131.7 mmHg, among all participants (sd = 17.8). The average serum cholesterol was 247.4 mg/dl (sd = 52) with fasting blood sugar < 120 mg/dl. Most of the resting electrocardiographic results have ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV). The average maximum heart rate achieved was 149.6 (sd = 23). Of the 279 patients, 97 had exercise induced angina and 200 did not. The average ST depression induced by exercise relative to rest was 1.1 (sd = 1.2). In the peak exercise ST segment, 200 of the patients have upsloping waves, 97 have flat waves, and 0 have downsloping waves. There are 174 patients with 0 major vessels colored by fluoroscopy, 65 with 1 vessel, 38 with 2 vessels, and 20 with 3 vessels. The thallium stress test showed that 164 of the patients had normal scores, while 115 had a reversible defect and 118 had a fixed defect.

To visualize variable interaction, variables were into distinct categories. Variable interaction was key determinant in understanding which variables had the strongest relationship with heart disease. Each variable was grouped into partitions of patients with heart disease and patients without heart disease. This yielded a mean for each variable, stratified across the two groups. The mean of the patients with heart condition was divided by the mean of patients without heart condition to find a ratio change. This ratio was driven by the "heart condition" variable. Ratios most divergent from 1.0 suggested which variables—based on this dataset—were most closely associated with cardiovascular disease.

```
## # A tibble: 13 × 4
##    predictor   heart_condition no_heart_condition ratio_change
##    <chr>                 <dbl>              <dbl>        <dbl>
##  1 ca                    1.15               0.275        4.17
##  2 exang                 0.540              0.144        3.76
##  3 thal                  1.37               0.375        3.66
##  4 oldpeak               1.59               0.599        2.65
##  5 slope                 0.825              0.412        2.00
##  6 sex                   0.818              0.556        1.47
##  7 chest pain            2.58               1.79         1.44
##  8 restecg               1.18               0.844        1.39
##  9 age                  56.8               52.6          1.08
## 10 blood pres          135.               129.           1.04
## 11 cholesterol         252.               243.           1.03
## 12 fbs                   0.146              0.144        1.02
## 13 thalach             139.               159.           0.877
```

Figure 1: Results of the ratio change analysis. A greater ratio change indicates variables that have potentially significant differences between patients with heart disease and those that do not have the disease.

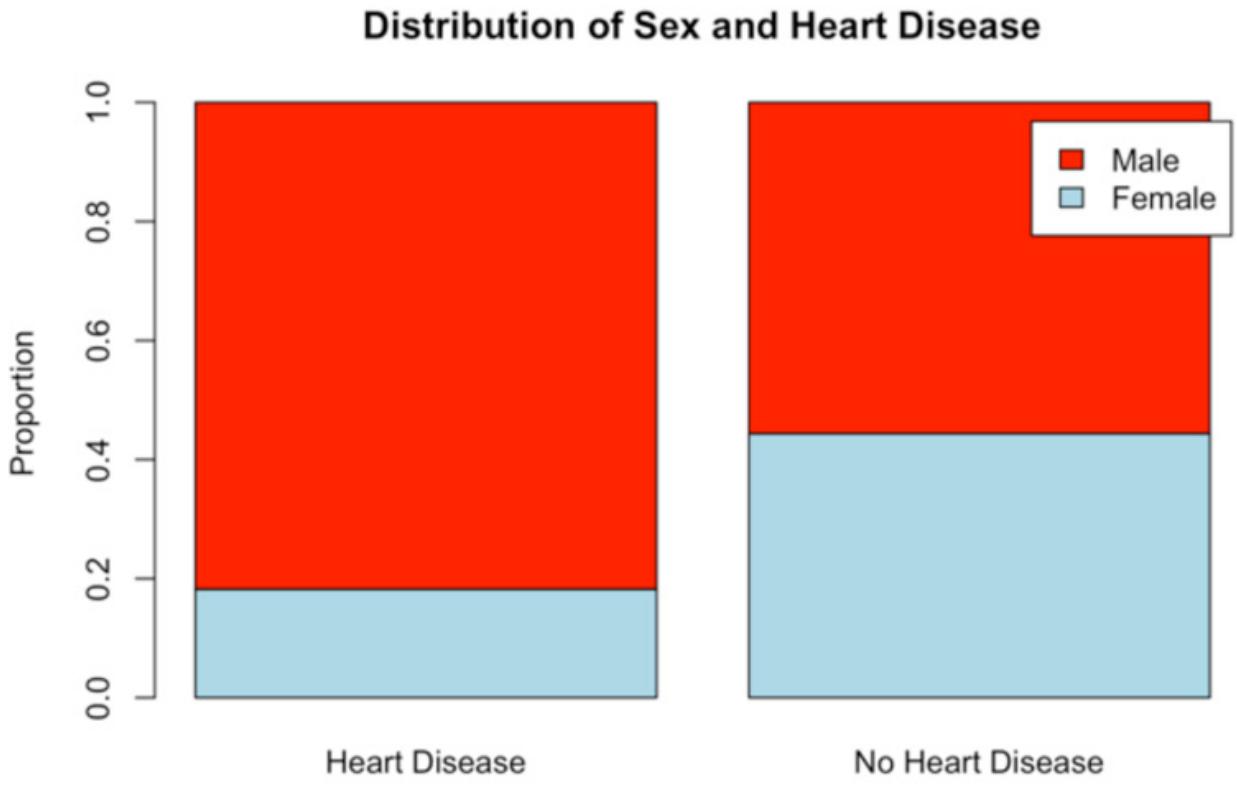


Figure 2: The distribution of the data between patients who endorse heart disease and those that do not, stratified by sex. The visualization shows that men may be more likely to have heart disease than women, as the "Heart Disease" group is comprised of a greater percentage of men than the "No Heart Disease" group.

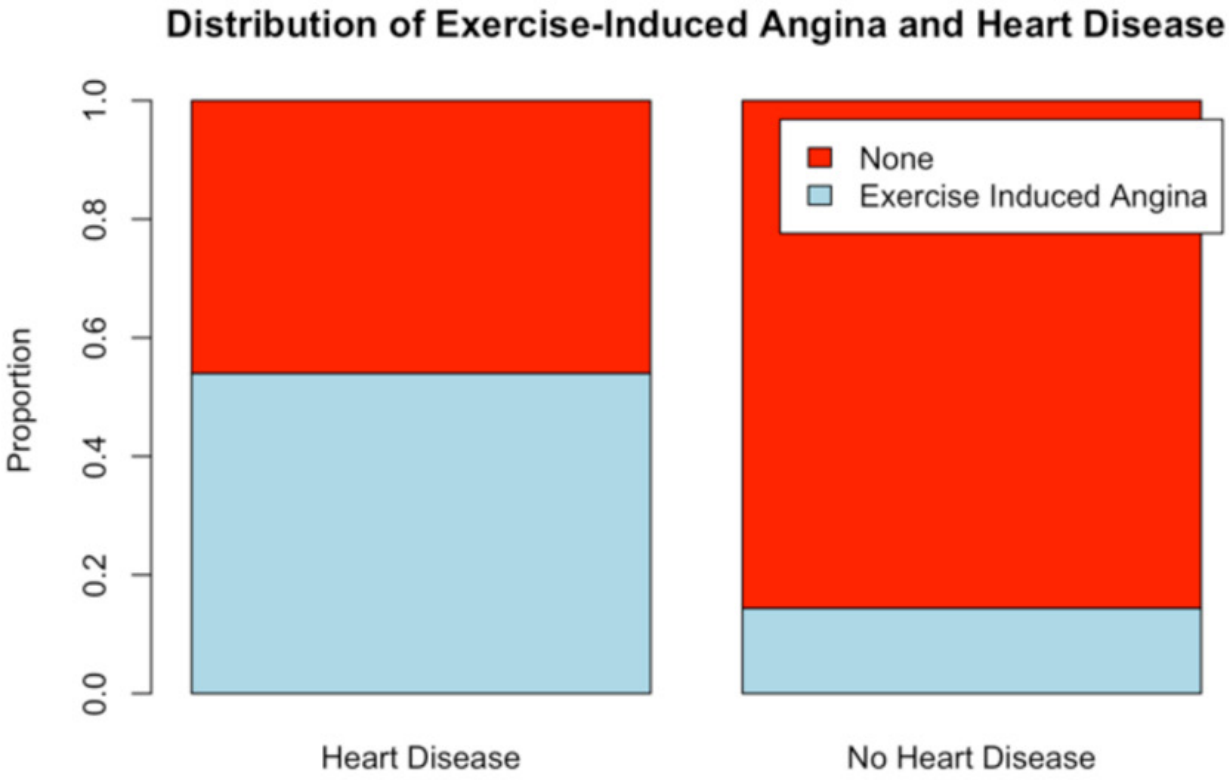


Figure 3: Distribution of the data between patients who have heart disease and those that do not, stratified by patient- reported experience of exercise-induced angina. The bar chart shows that amongst patients with heart disease, there exists a greater proportion of patients with exercise induced angina relative to those that are not affected by a form of cardiovascular disease.

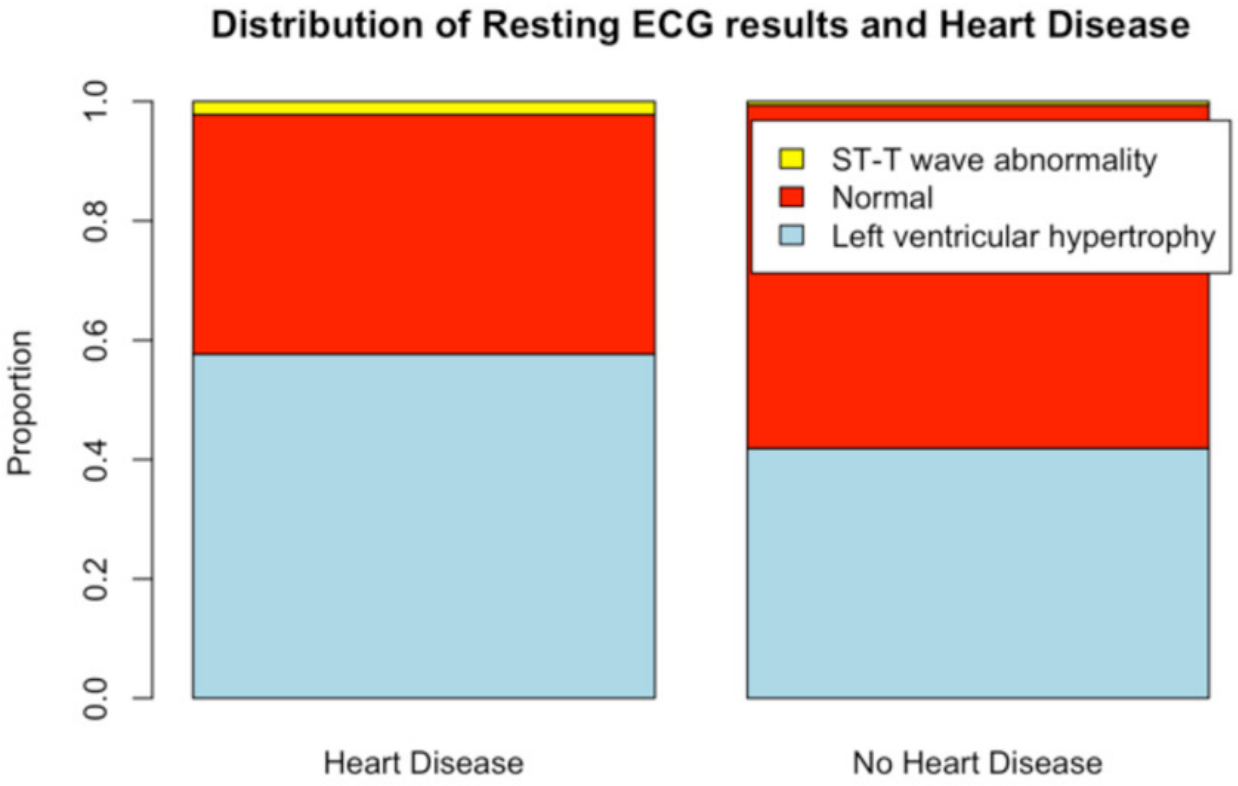


Figure 4: Distribution of the data between patients who have heart disease and those that do not, stratified by their resting electrocardiographic results. The data suggests that patients with heart disease have a slightly higher proportion of patients that also have left ventricular hypertrophy. Overall, there are very few patients in this study with ST-T wave abnormality. However, there are more patients with this abnormality that have heart disease than those who do not.

The most significant differences were found in three variables: exercise induced angina, resting electrocardiographic results, and gender. These variables are expected to be strong predictors of heart disease. The compilation of these three variables measures how much stress is put on the heart, controlling for differences in sex. Other variables, such as number of vessels colored by fluoroscopy and thallium stress test score, also showed significant results. However, these variables are more of a proxy to heart disease rather than a predictor of it. *Therefore, this analysis will explore the efficacy of utilizing exercise induced angina (exang), resting electrocardiographic results (restecg), and gender as valuable predictors of heart disease*.

The final component of this analysis included a logistic regression model fit to assess the role of sex, exercise induced angina, and resting electrocardiographic results in predicting heart disease. The condition variable for heart disease has a binary code (0 = no disease, 1 = disease). Thus, the model derived from covariates of this refined data set suggests the likelihood of disease, with a predicted condition value closer to 1 indicating a higher risk.

A chi-squared test is used to determine a relationship between each of the three indicated variables and the condition variable. This step is crucial in demonstrating that there is a significant association between the variables that could be predictors of heart disease and the "heart disease" variable itself. The data meet all necessary assumptions for chi-square testing. First, the assumption of independence is met as each patient is independent from another. Yet the second assumption (that the contingency table has values greater than 10) is not met for resting ECG results because there are not more than 10 values in that cell. Therefore, a Fisher test is used to examine the difference between an expected value and the observed value. The Fisher's Exact test is a statistical test that determines if there are nonrandom associations between two categorical variables. For example, when assessing exercise induced angina, the Fisher Exact test measures the difference between the expected number of patients who have heart disease and exercise-induced angina. Using holistic probabilities, the test compares the observed number of patients with both these conditions. A large difference between the expected and observed data suggests that there may be an association between the variables. This level of significance is quantified by a p-value. Confidence intervals are also be calculated for the three variables as further assessment of the null hypothesis.

Lastly, we develop a preliminary predictive model of whether a patient is expected to have a heart condition, based on their resting electrocardiographic results, self-reported exercise-induced angina status, and sex. A logistic regression is run on the three predicting variables as an estimate of whether the patient has heart disease or not. We have chosen to perform a logistic regression rather than a linear regression, as the logistic model allows us to assess the odds ratio for each individual variable in a simple way patients might understand. Rather than a collection of abstract coefficients. each value in the model represents the "increase is odds" of developing a heart condition, given that the patient endorses the variable of interest. Based on the three predicting variables, a log odds model is fit, and the odds ratios are calculated to assess the predictive values of the three variables. By taking the natural log of the odd (the probability), the product is "logit" of the probability, in other words, the log odds. This is useful because although the probability has restrictions between 0 and 1, the log odds can exist as a value anywhere between $-\infty$ to ∞. The log odds can be fit logistic regression model to yield the probability that a given patient has one of the indicated predictors and heart disease. We conducted hypothesis testing for each slope calculated in the log odds model, with the null being a log odds ratio of 1 and $\hat{\beta}$ of 0. This testing confirms the association found from the previous tests in this study. The confidence intervals are also taken for the odds ratio to demonstrate the likely range for that value, with 95% confidence. This test is important in our analysis, as we are interest in whether the value 1 exists within the calculated range. If βj is 1, that means the odds are equal and the results are not significant. If the calculations fall outside of 1 entirely, there is evidence of significant results. The model with the best adjusted r-squared value is then used to generate predictions.

Results

Testing Association of Categorical Variables of Interest

To assess the presence of a relationship between each of the predictors of interest and heart condition, a chi-squared test is conducted for each covariate.

```
##
##                Female Male
##   Condition        25  112
##   No Condition     71   89
```

```
##
##                Ex. Induced Angina No Exang
##   Condition                    74       63
##   No Condition                 23      137
```

```
##
##                Highly abnormal ECG Moderate ECG abnormality No ECG abnormality
##   Condition                     79                         3                  55
##   No Condition                  67                         1                  92
```

Figure 5: Table displaying the contingency tables for each categorical test for independence. The sex, exercise-induced angina, and rest ECG variables are all compared to the condition variable.

For the entirety of the study, patients are independent of one another, and these cells are mutually exclusive (given the distinct thresholds of the predictors). Thus, the chi square assumptions are met for the "sex" and "exang" contingency tables, as $n > 10$ in each cell. However, this assumption is failed in the "RestECG" contingency table. The Fisher's Exact test is non- parametric. A non-parametric test is a statistical test that uses a non-parametric statistical model. Such a model makes fewer assumptions than a parametric one regarding the distribution of the parameter of interest. Though non-parametric tests have less statistical power, they are valuable in that they can be used when other conditions are not met. Thus, the non-parametric Fisher's Exact test is used here.

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table_sex
## X-squared = 21.852, df = 1, p-value = 2.946e-06
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table_exang
## X-squared = 50.943, df = 1, p-value = 9.511e-13
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  table_ecg
## p-value = 0.004834
## alternative hypothesis: two.sided
```

Figure 6: Results for each Chi Square test and the non-parametric Fisher test. The associated p-value for each result is displayed.

The chi-square test for independence of condition and sex yields a test statistic X-squared = 21.85, which under the null hypothesis has a chi-square distribution with 1 degree of freedom. This test resulted in a p- value of 2.946e-06. Since p < alpha (2.946e-06 < 0.05), the null hypothesis that there is no association between heart condition and sex is rejected. This analysis suggests an association between heart condition and sex of the patient. More specifically, this dataset—in support of previous research—further suggests that being male is associated with an increased risk of heart disease.

The second chi-square test for independence of condition and exercise induced angina yields a test statistic X-squared = 50.94, which under the null hypothesis has a chi-square distribution with 1 degree of freedom. This test resulted in a p-value of 9.511e-13. Since p < alpha (9.511e-13 < 0.05), the null hypothesis that there is no association between heart condition and exercise induced angina is rejected. There is sufficient evidence to suggest that there is an association between heart condition and whether a patient has exercise induced angina. Based on the data, it is likely that having exercise induced angina is associated with an increased the risk of heart disease.

The Fisher test for independence of condition and resting ECG yields a p-value of 0.00483 for the count data above. Since p < alpha 0.00483 < 0.05), the null hypothesis that there is no association between heart condition and a patient's resting ECG is rejected. Though this is a non-parametric test, there is still sufficient evidence to suggest that there is indeed an association between the level of abnormality of a patient's resting ECG and heart condition. The data suggest that abnormal resting ECGs are associated with a higher risk of heart condition.

Regression Analysis

To develop a preliminary predictive model for a patient's expected odds of heart condition, a log odds model will be created. Collectively, the predictors will yield the probability of a patient having a heart condition, given their sex, whether they have exercise-induced angina, and the level of abnormality of their resting ECG results. To fit a log odds model, several

assumptions must first be met. The probability of heart condition must be independent in for each patient. Even though the patients belong to the same city because heart disease is a noncontagious condition, this assumption of independence is met. Heart condition is also a Bernouilli variable because condition of an individual is considered a "success". Lastly, the effect of the predictors must be linearly related to the log odds. Plotting the residuals to the fitted values, there are two lines with constant slopes, so this assumption is met.

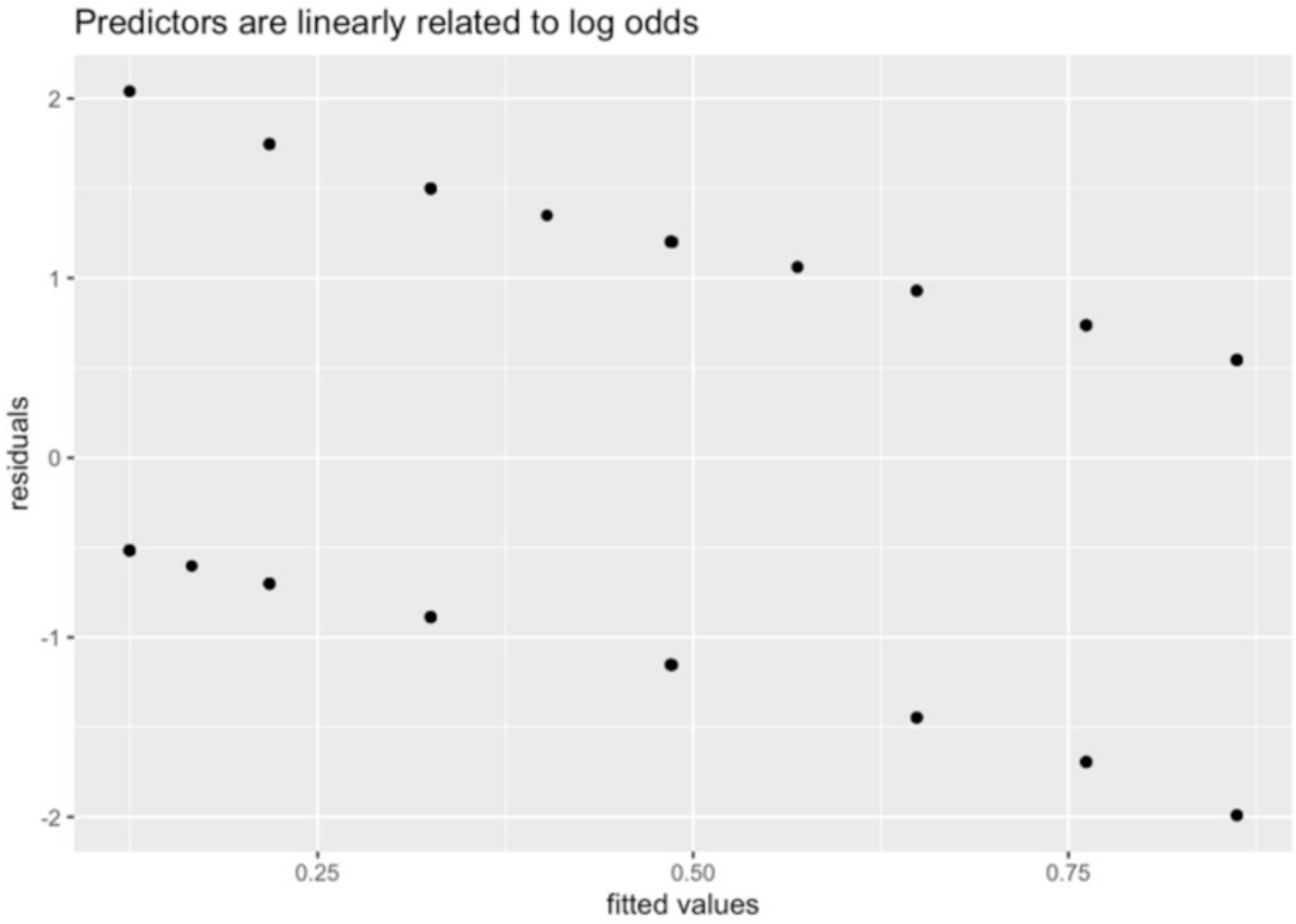


Figure 7: Plot of the residuals against the fitted values for the regression model. This plot illustrates that the assumption of linearity is met.

With these assumptions met, a log odds model can be constructed.

```
## # A tibble: 4 × 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)   -1.95      0.314     -6.20 5.60e-10
## 2 exang          1.89      0.296      6.40 1.57e-10
## 3 sex            1.22      0.301      4.05 5.08e- 5
## 4 restecg        0.335     0.135      2.48 1.32e- 2
```

Figure 8: Summary of the log odds regression model. The table features the intercept, the slope associated with each variable, and the standard error of that slope estimate along with the associated test statistic and p value.

This yields the following regression model:
Predicted log odds of condition = -1.948 + 1.89(exang) + 1.218(sex) + 0.335(restecg)

Each slope in this model provides a value to explain the degree to which the predictor contributes to the likelihood of condition. Exponentiation of these variables explains the exact effect that the predictor has on the log odds ratio.

Exponentiation of the "exercise-induced angina" slope returns a value of 6.63. When all other predictors are held constant, a patient that has exercise induced angina is 6.63 times more likely to have heart condition. This slope was the most significant, with an associated p-value of approximately 1.57e-10. Since p is much less than the significant level of 0.05, the null hypothesis that the log odds ratio is 1 is rejected, and there is sufficient evidence to claim that there is a relationship between exercise induced angina and the likelihood that a patient has heart disease.

Exponentiation of the "sex" slope returns a value of 3.38. When all other predictors are held constant, a patient that is a male is 3.38 times more likely to have heart condition. This slope was significant, with an associated p-value of approximately 5.08e-5. Since p is less than the significant level of 0.05, the null hypothesis that the log odds ratio is 1 is rejected, and there is sufficient evidence to claim that there is a relationship between sex and the likelihood that a patient has heart disease, with the data indicating males are at higher risk.

Exponentiation of the "resting ECG" slope returns a value of 1.40. When all other predictors are held constant, a one unit increase in the degree of abnormality of a patients' ECG is associated with that patient being 1.4 times more likely to have a heart condition. This slope was significant, with an associated p-value of approximately 0.013. Since p is less than the significant level of 0.05, the null hypothesis that the log odds ratio is 1 is rejected, and there is sufficient evidence to claim that there is a relationship between resting ECG and the likelihood that a patient has heart disease, with the data indicating patients with abnormal resting ECG results are at a higher risk of heart condition.

To further evaluate the significance of these log odds ratios, 95% confidence intervals (CIs) can be calculated for each. First, a 95% CI is constructed for the odds ratio having heart condition comparing males to females.

The 95% CI for the odds ratio of heart condition corresponding to the patient being male (as opposed to female) is (1.65, 6.93). This means that if samples were taken from the population and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the true population log odds ratio for heart condition. Conversely, 5% of such constructed intervals would not encompass the log odds ratio. A log odds ratio that is not equal to 1 indicates that there is an association between sex and heart condition. Thus, for there to be a significant association, the CI to would be expected to not encompass the value 1. Since 1 falls outside the bounds of this confidence interval there is sufficient evidence to suggest that being a male has an association with increased odds of heart disease. Secondly, the association between exercised-induced angina and heart condition is explored.

The 95% CI for the odds ratio of heart condition corresponding to the patient having exercise induced angina is (2.22, 19.87). This means that if samples were taken from the population and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the true population log odds ratio for heart condition. Conversely, 5% of such constructed intervals would not encompass the log odds ratio. Since 1 falls outside the bounds of this confidence interval there is sufficient evidence to suggest that having exercise induced angina has an association with increased odds of heart disease.

Lastly, the relationship between resting ECG results and heart condition is explored.

This CI is (1.28, 1.58), and can be interpreted the same way as the former two. Since 1 falls outside the bounds of this confidence interval there is sufficient evidence to suggest that having an abnormal resting ECG has an association with increased odds of heart disease.

Using Regression Model to Generate Predictions

Thus far, we have demonstrated that there is a significant association between each variable of interest and heart condition. The regression model is shown has significance with each predictor. To explore practicality of using this model to predict the

likelihood of heart condition, consider the scenario of a patient that is a male, with exercise-induced angina, but a perfectly normal resting ECG.

Statistics for hypothetical male patient with normal resting ECG and exercise induced angina

The predicted log odds are approximately 1.163. The parameter of interest is the probability of heart disease, p. It is known that the log odds is log(p/(1-p)). Solving this equation yields a probability p of 76.2%. Therefore, for the patient described above, there is a 76.2% chance that he has a heart condition. This indicates that being a male with exercised induced angina places a patient at a high risk for heart disease. Using the three variables of interest, this model can provide a preliminary and/or suggestive sense of the probability of heart disease for a patient.

Discussion

This analysis of the Cleveland-UCI Heart Disease data set explored how exercise induced angina (exang), resting electrocardiographic results (restecg), and gender can be used in predicting heart disease. A preliminary analysis of the data was conducted using bar plots. Each predictive variable was visualized separately. All three bar plots suggested a potential association between the given predictive variable and the patient's diagnosis status (i.e.., has heart disease or does not have heart disease). A comprehensive analysis of these three predictive variables supported this initial finding.

For sex and exercise-induced angina, the assumptions for a chi-square test were met. However, the assumptions were not met for resting ECG. Thus, a Fisher's Exact test was run instead. A chi-square test demonstrated that there was statistically significant evidence that an association between heart disease and sex is present (p-value 2.946e-06). A chi-square test also demonstrated that there was statistically significant evidence that an association between heart disease and whether or not a patient experienced exercised-induced angina is present (p-value 9.511e-13). A Fisher's Exact test additionally demonstrated that there was statistically significant evidence that an association between heart disease and the resting ECG of a patient is present (p-value 0.0048). A significant association was found between heart condition and all three predictive variables independently of each other. To better quantify this association, a logistics regression model was fitted to the data. The assumptions were met for a logistic regression model using the three predictive factors. The resulting model was "yhat= -1.948 + 1.89(exang) + 1.218(sex) + 0.335(restecg)". The corresponding p-values for all three slopes showed a significant association between the predictors and heart disease. Similarly, the 95% confidence intervals for the odds ratio of each of the predictors provides evidence of a significant association between each predictor and heart condition.

Overall, the data analysis showed that angina (exang), resting electrocardiographic results (restecg), and gender may be useful in predicting heart disease since each has a statistically significant association with a heart condition. The logistic regression model shows that of the three variables explored, exercise-induced angina is expected to be the best predictor of heart condition since it has the highest odds ratio.

The finding that all three variables can be used to predict heart disease is an important preliminary discovery that can play a role in furthering the understanding of clinical diagnostic tools. Since being male is linked to a higher risk of heart disease, this information can be used to inform clinical practices. Extra precautions can be taken to detect potential heart disease in male patients. The presence of exercised-induced angina might be a helpful diagnostic factor since it does not require any procedures or medical evaluations and can be determined by the patient prior to a clinical appointment. Patients who suspect heart disease or are at risk for heart disease can assess themselves during exercise for the presence of angina. Lastly, an ECG can be used to further screen patients suspected of having heart disease. Although this is a procedure that requires time and resources to utilize, it is still non-invasive and relatively easy to administer. Together, these three variables offer preliminary predictive power in determining individuals that are or may be at risk for a heart condition.

The data set used is a well-documented data matrix, with dozens of citations in high impact academic publications. Thus, the data is both reliable and generalizable to the U.S. public, with variables

being collected from five different clinical locations across the nation exhibiting racial and ethnic diversity. A total of 303 patients were enrolled in the study. For 297 patients, there was complete data for all three predictive variables used in this study. These patients were selected for our data analysis. Since the data was gathered in U.S. hospitals, the results cannot necessarily be generalized to patients in other countries or contexts. One must proceed with caution when using these results to inform practices in other countries or amongst different patient demographics than the one from this data set. It is important to explore the role of culture in these conclusions. One way to proceed in analyzing the strongly predicting factors for a heart condition of a patient in another country is following this study's methods of comparing the variables and running tests with data collected from that respective country. Within the data collection for this Cleveland Clinic data set, data from Hungarian Institute of Cardiology (Budapest, Hungary), University Hospital (Zurich, Switzerland), and University Hospital (Basel, Switzerland) can be compared to the Cleveland Clinic results for further analysis. Another limitation of this data set is its age. The data was released in 1988 and is thus between 30 and 40 years old. Shifting trends in public health and health policy may mean that the results from this data analysis do not fit the current population as well. Further research could repeat this methodology on a more recent data. Finally, it is important to keep in mind the complexity of heart disease and take caution in not "over-fitting" with this model. This model is simply a preliminary predictor and should be used as a tool in concert with a detailed understanding of the individual patient and their comprehensive past medical history.

Further research can also be done to better understand how to apply the results found in this study. A great starting point would be to utilize this model on a different test data set. Many have been made available by the CDC and NIH and are accessible to the public. Additional analysis may be performed to assess whether the predictive power of these three variables is consistent across differing levels and categories of other variables such as age, race, or daily stress levels. If this analysis were to be conducted again, potential interactions between the chosen variables and new variables could also be explored. Possible interaction terms could include resting ECG and resting blood pressure as well as ST depression induced by exercise relative to rest and maximum heart rate achieved. Any associations would help to better understand how to use sex, exercise-induced angina, and resting ECG as predictive variables for heart condition. Furthermore, standard predictiveness matrices could be calculated to provide greater insight in a future publication. Utilizing standard machine learning methods (such as CART or random forest), the model could be expanded to be more than a logistic regression. Overall, expanding on the predictive tools and assessing the power of accuracy of this model will be important steps in progressing this research, as well as the care of patients at risk for or with heart conditions.

Conclusion

The results of this study provide preliminary indications for predictors of a heart condition. Given that the variables identified in this research do not require laboratory blood work or specialty testing, this could present potentially relevant and easily accessible factors to be implemented into future models of diagnosis in primary care clinics and tools for cardiology specialist referrals. Most of the existing models, such as the 10-year ACSVD risk calculations, tend to rely on values from serum lab and heart and extensive medical history input (Murphy et al., 2018). As medicine becomes more specialized, communication between providers is extremely important. This research may be a useful underpinning for the development of future models that could ensure accurate referrals, bridging communicative gaps between primary care providers and cardiologist. Ultimately, it is proposed that future models will be able to improve the efficacy and accessibility. s.

References

1. Blesso, C. N., & Fernandez, M. L. (2018). Dietary Cholesterol, Serum Lipids, and Heart Disease: Are Eggs Working for or Against You?. *Nutrients*, 10(4), 426. https://doi.org/10.3390/nu10040426 (https://doi.org/10.3390/nu10040426)
2. Cherngs. (2020, March 29). Heart disease CLEVELAND UCI. Retrieved April 27, 2021,

from https://www.kaggle.com/cherngs/heart-disease-cleveland-uci (https://www.kaggle.com/cherngs/heart- disease-cleveland-uci)

3. De Bacquer, D., De Backer, G., Kornitzer, M., & Blackburn, H. (1998). Prognostic value of ECG findings for Total, cardiovascular disease, and coronary heart disease death in men and women. *Heart*, 80(6), 570-577. doi:10.1136/hrt.80.6.570 (doi:10.1136/hrt.80.6.570)
4. Franklin, S. S., Larson, M. G., Khan, S. A., Wong, N. D., Leip, E. P., Kannel, W. B., & Levy, D. (2001). Does the relation of blood pressure to coronary heart disease risk change With Aging? *Circulation*, 103(9), 1245-1249. doi:10.1161/01.cir.103.9.1245 (doi:10.1161/01.cir.103.9.1245)
5. Heart disease. (2021, January 19). Retrieved April 26, 2021, from https://www.cdc.gov/heartdisease/index.htm (https://www.cdc.gov/heartdisease/index.htm)
6. Kannel, W. B., Gordon, T., & Schwartz, M. J. (1971). Systolic versus diastolic blood pressure and risk of coronary heart disease. *The American Journal of Cardiology,* 27(4), 335-346. doi:10.1016/0002- 9149(71)90428-0 (doi:10.1016/0002-9149(71)90428-0)
7. Libby P, Theroux P. Pathophysiology of coronary artery disease. *Circulation*. 2005 Jun 28;111(25):3481-8. doi: 10.1161/CIRCULATIONAHA.105.537878. PMID: 15983262.
8. Marateb, H. R., & Goudarzi, S. (2015). A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system. *Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences*, 20(3), 214–223.
9. Milner, K. A., Funk, M., Richards, S., Wilmes, R. M., Vaccarino, V., & Krumholz, H. M. (1999). Gender differences in symptom presentation associated with coronary heart disease. *The American Journal of Cardiology*, 84(4), 396-399. doi:10.1016/s0002-9149(99)00322-7 (doi:10.1016/s0002-9149(99)00322-7)
10. Murphy SL, Xu JQ, Kochanek KD, Arias E. Mortality in the United States, 2017. *NCHS Data Brief*, no 328. Hyattsville, MD: National Center for Health Statistics. 2018.
11. WIENER, L., DWYER, E. M., & COX, J. W. (1968). Left ventricular hemodynamics in exercise-induced angina pectoris. *Circulation*, 38(2), 240-249. doi:10.1161/01.cir.38.2.240 (doi:10.1161/01.cir.38.2.240)

The Blue Devil Buddies Program: Preliminary Evaluation and Suggestions for Future Research in Support of Undergraduate Student Wellbeing at Duke University

Ishika Gupta, Nicole Izquierdo, Kiran Sundar, Victoria Wang

Volume 1, Issue 1

The Blue Devil Buddies Program: Preliminary Evaluation and Suggestions for Future Research in Support of Undergraduate Student Wellbeing at Duke University

Ishika Gupta, Nicole Izquierdo, Kiran Sundar, Victoria Wang

Department of Psycology, Duke University
https://doi.org/10.55894/dv1.13

Abstract

Duke University recently established the Blue Devil Buddies (BDB) program, a two- semester peer mentorship that pairs first-year students with upperclassmen to facilitate a smooth transition from high school to college. This study examined the efficacy of the program by measuring the psychological well-being and sense of belonging of students before and after the program. Although the initial hypotheses were not supported, there was a significant interaction effect of participation in the BDB program on psychological well-being, $F(1, 54) = 3.116$, $p = 0.083$, $\eta 2 = 0.043$, and sense of belonging, $F(1, 54) = 8.888$, $p = 0.004$, $\eta 2 = 0.091$. The semester of participation (either Fall 2020 or Fall 2021) did have a significant effect on psychological well-being scores, $F(1, 36) = 4.153$, $p = 0.049$, $\eta 2 = 0.081$; however, there was no significant interaction effect of semester and time on psychological well-being, $F(1, 36) = 2.263$, $p = 0.141$, $\eta 2 = 0.012$. Although there were no significant effects of semester on sense of belonging, $F(1, 36) = 0.811$, $p = 0.374$, $\eta 2 = 0.012$, there was a significant interaction effect of semester and time on post-program sense of belonging scores, $F(1, 36) = 6.435$, $p = 0.016$, $\eta 2 = 0.059$. The lack of effects indicates that a subsequent study with a much larger sample size must be conducted before any definitive conclusions about the program can be drawn. Nonetheless, the results hint at the possibility that the BDB program may attract students who are already thriving socially and mentally. Thus, this study lays the groundwork for future studies to investigate whether the program effectively identifies and supports the students who need peer mentorship the most.

Introduction

Research in clinical psychology shows that mental illness often develops during the first year of college (Pedrelli et al., 2015). Indeed, the transition from high school to university comes with a host of major changes to daily routine, social life, and academic rigor. These disruptions can make first-year students particularly vulnerable to mental health challenges. Past research shows that peer mentorship can protect individuals against potential harms to mental well-being by offering direct support and a sense of community, two social factors linked directly to positive mental health (Shalaby & Agyapong, 2020).

Duke University's Student Government (DSG) recently implemented a mentorship program for first-year students called the Blue Devil Buddies (BDB) program. It was started in 2019 to help incoming freshmen match with an upperclassman with similar academic and personal interests. Its purpose is to alleviate the difficulties that come with the rapid transition from high school to college. However, because this is a new program, there is little data collected on its effectiveness. To conduct a primary analysis of the program and identify areas for potential improvement, this study focused on collecting data from upperclassmen mentors, first- year mentees, and other Duke students who did not participate in the BDB program. The primary research question in this investigation is, "How effective has the Blue Devil Buddies program been in improving mental and social well-being among Duke undergraduate students?"

Prior research related to adolescent mentorship has demonstrated largely promising results. In one study, peer mentorship for first-year college women was shown to protect against adverse physical outcomes, such as unhealthy increases in body fat percentage and decreases in cardiovascular fitness (Wright et al., 2018). At the University of Pittsburgh, peer mentorship demonstrated a clear positive impact on first-year students through emotional guidance. Moreover, the students in this study who had proactive and engaged mentors were more likely to succeed in their academic, family, and personal life (Budny, Paul, & Bon, 2006). Other researchers in the field found that peer mentorship improved the academic performance and self- efficacy of college students enrolled in an introductory psychology class: students who received peer mentoring consistently achieved higher exam grades, and therefore a higher final course grade, than their non-mentored counterparts (Asgari & Carter, 2016). Notably, student responses to post-mentorship surveys indicated that having a peer mentor increased their confidence and motivations to succeed. These findings suggest a link between peer mentorship and overall well- being. However, these benefits were dependent on the mentor's behavior. Similarly, Holt and Fifer (2016) observed that the attachment styles and self-efficacy of mentors affected the perceived quality of the mentorship on both sides.

Although mentorship programs have been examined before, few studies examine the experience of both mentors and mentees in a program targeting first-year college students. One potential lack of research identified in the literature is a lack of measurement of programs effects on social well-being (e.g., how students perceive their social support or isolation after the program). A second potential gap recognized in the literature is a comparison of the differences between mentee and mentor experiences. It is important to separately analyze these experiences because it is likely that mentors and mentees experience different social and mental outcomes as a result of the program. A third important gap identified is that previous studies often do not compare baseline data for students who participate in the mentorship programs to students who do not participate; such a comparison could yield data on how students who join these types of programs may systematically differ from students who do not. In an effort to begin to address these three gaps, the present study measures psychological and social well-being before and after the mentorship program, compares survey data between mentors and mentee experience, and analyzes survey data between program participants and non-participants.

It was hypothesized that being in the program in any capacity, either as a mentee or mentor, resulted in significantly higher psychological and social well-being when compared to non-participants. Additionally, within the program, it was hypothesized that mentees would show significantly greater changes in their sense of belonging score (social well-being) compared to mentors. Mentees would most likely benefit more from the BDB program on the sense of belonging scale, especially because mentors are already established Duke students while mentees are transitioning from a high school environment to Duke's campus. If this improvement in sense of belonging prior to the program and after the program did occur in mentees, it could imply that their transition from high school to college was facilitated by the BDB program.

It is important to note that not every mentee and mentor relationship is equal. Some students may feel like they found their perfect match, while others may not perceive their mentorship to be beneficial. As illustrated in the previous literature, it is possible that the mentor/mentee relationship could mediate how much a given student benefited from the program. Specifically, we hypothesize that the higher a student's perceived quality of their mentor/mentee relationship, the more likely they were to benefit from the program, demonstrated by a higher score on the psychological and social well-being scales.

Another important factor in the program was the COVID-19 pandemic. Due to this, in-person interactions were limited or non-existent, so some mentor/mentee pairings met mainly through Zoom or FaceTime. It was hypothesized that the greater the proportion of digital interaction, the lower the students would score on both well-being scales. It was also hypothesized that there would be an interaction between the quality of interaction and proportion of digital interaction factors.

Below are the hypotheses:
Hypothesis 1: People in the program with any role (mentor or mentee) will score higher on both the psychological well-being and sense of belonging scales.
Hypothesis 2: Mentees will score higher than mentors on the sense of belonging scale, but not on the psychological well-being scale.
Hypothesis 3: The greater the perceived quality of interaction between mentor and mentee, the higher the scores on both scales.
Hypothesis 4: The greater the proportion of digital interaction between mentor and mentee, the lower the scores on both scales
Hypothesis 5: An interaction will occur between perceived quality of interaction and proportion of digital interaction.

Regardless of the results, this study will provide DSG with preliminary analysis of their program. The results have the chance to inform further research on the program, eventually leading to insights on the program's current efficacy, as well as areas for improvement. It is important that programs like this are evaluated on their efficacy to explore whether they are reaching their goals, to determine if certain situations or populations experience different outcomes, and to better inform future iterations and related programs. Finally, this analysis will give a pilot understanding of how the program is living up to it goal of "foster[ing] community through conversations ranging from course selection to campus life."

Methods

Participants

Duke undergraduate students were sampled as the population for this study. Due to limited resources, a convenience sample was used. The Qualtrics survey link was sent through Facebook groups (Class of 2025, Class of 2024, Class of 2023, Class of 2022, etc.) and group chats (Mi Gente, ASA, SOCA, University Scholars, etc.). There were 98 total responses to the survey; however, 42 were excluded because they were incomplete. Analysis of survey response were conducted on the remaining 56. The only exclusionary criteria was anyone under 18 years of age. If a participant listed their age as 17 or younger, their data was excluded. The 56 responses used consisted of a range of class years (first-years, sophomores, juniors, and seniors). The present study did not ask demographic questions (such as race or gender), so it is unclear how demographics could have affected survey responses. As compensation for completing the survey, each participant was entered into a raffle for a chance to win $50 (either as a gift card for themselves or a donation to an organization of their choice).

Materials

The survey was delivered on electronically on Qualtrics. There was a total of 270 items, but participants only answered a portion of these depending on their responses to previous survey items. For instance, if a participant indicated that they were a mentor starting Fall 2020, then they were taken only to the questions that apply to their position. The survey took the participants in a range of approximately 5-15 minutes to complete. Participants could complete the survey on a computer, laptop, tablet, or phone; however, the survey interface was optimized for a computer or laptop.

To operationalize mental well-being, the study used an existing scale of measurement developed by researchers in the field: the Warwick-Edinburgh Mental Well-being Scale (Tennant et al., 2007). This scale measures positive affect, satisfaction in interpersonal relationships, and positive functioning. The questions in this scale include, "During the specified semester, how often did you experience the following?". Participants were asked to rate statements, such as "I felt optimistic about the future," "I felt useful," or "I felt relaxed," on a 5- point Likert scale from 1 being "none of the time" and 5 being "all of the time." Scores were computed as a sum of the participant's responses to all the questions. The scale was found to have high test-retest reliability after one week (0.83, $p < 0.01$) and good content validity as assessed by frequency of complete responses and the distribution of responses to each item (Tennant et al., 2007).

Social well-being was operationalized as sense of belonging through the revised Hoffman Sense of Belonging Scale (Morrow et al., 2002). This scale

has 26-items and measures perceived peer support, perceived classroom comfort, perceived isolation, and perceived faculty support. The survey used a revised version of this scale since perceived faculty support was not relevant for this study. To quantify the individual factors included in the scale (i.e., perceived peer support, perceived classroom comfort, perceived isolation), a score was calculated following the guidelines established by the original investigators. Each item was on a 5-point Likert scale from 1 being "completely untrue" to 5 being "completely true." Scores were computed as a sum of the participant's responses to all the questions. The type of prompts on this scale included: "I have met with peers outside of class to study for an exam," "I feel comfortable contributing to class discussions," and "It is difficult to meet other students at Duke." The scale was found to have high content validity and internal consistency within each factor (Morrow et al., 2002).

For quality of interaction, two existing scales originally developed for the Big Brothers Big Sisters community-mentorship programs were used: the Youth Strength of Relationship (YSoR) and the Mentor Strength of Relationship (MSoR). The YSoR was used for mentees and the MSoR for mentors. The YSoR and MSoR scales are brief surveys that capture both positive and negative aspects of a mentorship pairing. Both mentor and youth versions also have demonstrated good reliability for the total scores and associations with match length in a sample of BBBS community-based matches (Rhodes et al., 2016). For further information on the references and contents of the original scale, please see the Appendices section.

Most of the original survey items from the scales were revised to fit the study's context (Duke undergraduate students). Certain words and verb tenses were replaced and edited.

Procedure

A correlational design where the groups were already defined was utilized because the study was based on a recent program. Because the program was recent enough, there was a low chance of bias with participants not being randomly assigned. Therefore, the participants were not randomly assigned to groups/roles. The independent variables included the student's role in the BDB program (mentor, mentee, none), the proportion of the digital interaction with their mentor or mentee (low or high), and the quality of the interaction with their mentor or mentee (low or high, YSoR or MSoR). The dependent variables were sense of belonging and psychological well-being, which were operationalized using two scales (the revised Hoffman Sense of Belonging scale and the Warwick-Edinburgh Mental Well-being scale). Participants were given the choice to opt-in to the study and answer the questions or opt-out of the study and exclude themselves from the sample.

Data Analysis Plan

Two separate analyses were run across the three groups. For both, the analysis protected against alpha error-rate inflation by using the Bonferroni p-value as reference such that the Bonferroni p-value must be less than 0.05 for the effect to be considered significant.

For the first analysis, the data provided by Blue Devil Buddies mentors and mentees was used to run a three-way ANOVA on a 2 x 2 x 2 factorial design. First, an omnibus test was conducted to determine whether there were any significant effects overall. A post-hoc main effect tests was unnecessary because there were only two groups under study. However, if the omnibus test led to a rejection of the null hypothesis, post-hoc tests would be performed to determine whether there were any significant interaction effects between the variables. This determined whether any of the factors (e.g., mentor versus mentee, proportion of digital versus in-person interaction, perceived quality of interactions) differentially affect mental and social well-being among students.

For the second analysis, data from all three groups (i.e., mentors, mentees, and non- participants) were compared in three separate one-way ANOVAs. As before, an omnibus test was performed to determine whether the null hypothesis was rejected. If the results were significant, then post-hoc tests were conducted to identify the specific main effects. This offered some insight into the general efficacy of the program by revealing potential differences in the two measures of well-being between students who participated in some way and students who did not participate at all.

Results

The variables analyzed were scores on the psychological well-being pre-test, the sense of belonging pre-test, the psychological well-being post-test, the sense of belonging post-test, the quality of interaction, and the proportion of digital interaction. See Descriptive Tables 1-4.

The first hypothesis was that people in the program with any role (mentor or mentee) would score higher on both the psychological well-being and sense of belonging scales. Based on an ANOVA, there were no significant main effects of program participation on psychological well-being, $F(1, 54) = 0.142$, $p = 0.708$, $\eta 2 = 0.003$. See Figure 1. There were no significant main effects of program participation on sense of belonging, $F(1, 54) = 0.109$, $p = .743$, $\eta 2 = 0.002$. See Figure 2.

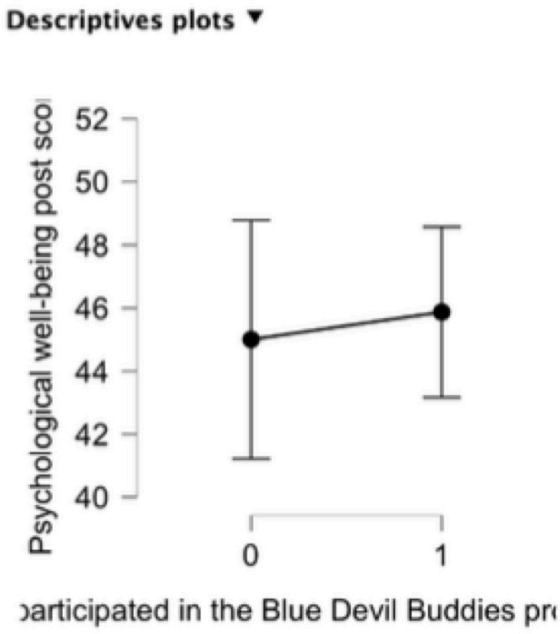


Figure 1: Participation in the program (0 = no & 1 = yes) & psychological well-being post scores

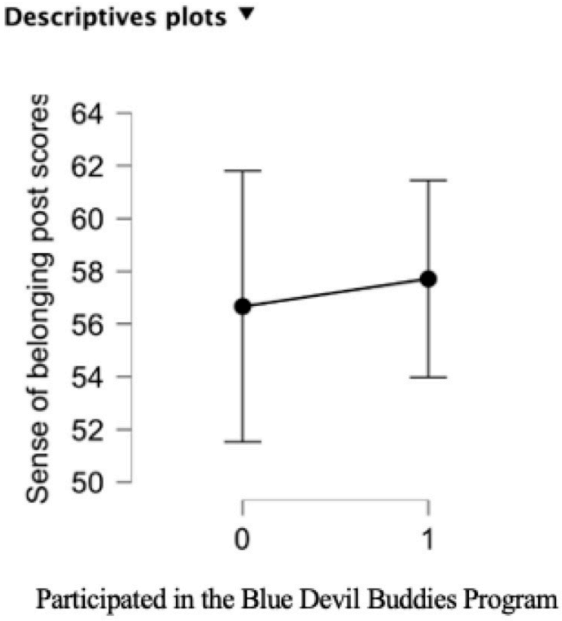


Figure 2: Participation in the program (0 = no & 1 = yes) & sense of belonging post scores

The second hypothesis was that mentees would score higher than mentors on the sense of belonging scale, but not on the psychological well-being scale. There were no significant main effects of role in the program on psychological well-being, $F(1, 36) = 1.081$, $p = .305$, $\eta 2 = 0.029$. See Figure 3. There were no significant main effects of role in the program on sense of belonging, $F(1, 36) = 1.939$, $p = 0.172$, $\eta 2 = 0.051$. See Figure 4.

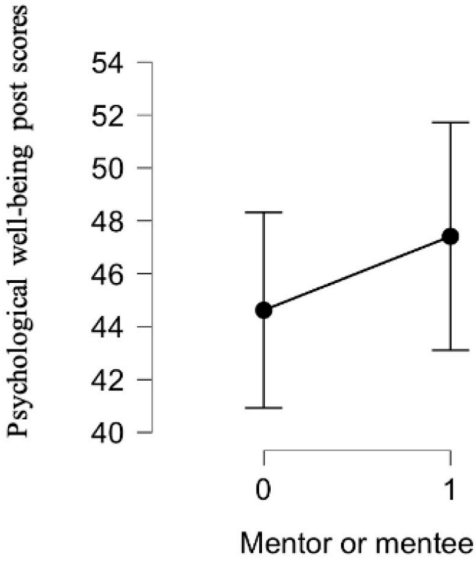


Figure 3: Role in the program (0 = mentee & 1 = mentor) & psychological well-being post scores

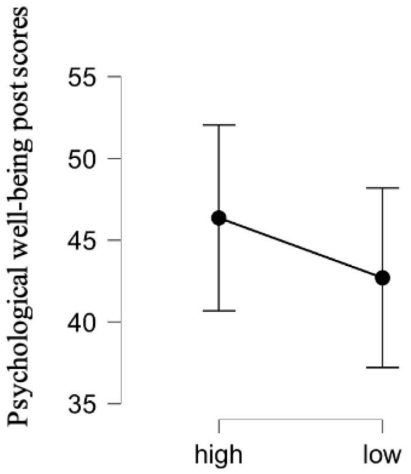


Figure 4: Role in the program (0 = mentee & 1 = mentor) & sense of belonging post scores

The third hypothesis was that the greater the perceived quality of interaction between mentor and mentee, the higher the scores would be on both scales. The survey measured quality of interaction as a continuous variable, but the analysis used a dummy coded version based on a median split between low and high to create a categorical variable. The median for mentees was 36, so scores below 36 were coded as low and scores equal to or above 36 were coded as high (≤ 35 = low & ≥ 36 = high). The median for mentors was 50, so scores below 50 were coded as low and scores equal to or above 50 were coded as high (≤ 49 = low & ≥ 50 = high). There were no significant main effects of quality of interaction for mentees in the program on psychological well-being, $F(1, 19) = 1.072$, $p = 0.314$, $\eta 2 = 0.053$. There were also no significant main effects of quality of interaction for mentors in the program on psychological well-being, $F(1, 15) = 0.492$, $p = 0.494$, $\eta 2 = 0.032$. There were no significant main effects of quality of interaction for mentees in the program on sense of belonging, $F(1, 19) = 0.638$, $p = 0.434$, $\eta 2 = 0.032$. There were no significant main effects of quality of interaction for mentors in the program on sense of belonging, $F(1, 15) = 0.530$, $p = 0.478$, $\eta 2 = 0.034$.

The fourth hypothesis was that the greater the proportion of virtual/digital interaction between mentor and mentee, the lower the scores on both scales. Since the median for proportion of virtual meetings was 3, 1-2 was considered low proportion of digital interaction and 3-5 as high proportion of digital interaction. There was no significant main effect of the proportion of digital interaction on psychological well-being, $F(4, 33) = 0.123$, $p = 0.973$, $\eta 2 = 0.015$. There was also no significant main effect of the proportion of digital interaction on sense of belonging, $F(4, 33) = 0.543$, $p = .705$, $\eta 2 = 0.062$.

The fifth hypothesis was that an interaction would occur between perceived quality of interaction and proportion of digital interaction on both scales. There were no significant main effects of quality of interaction for mentee's in the program on psychological well-being, $F(1, 19) = 1.072$, $p = 0.314$, $\eta 2 = 0.053$. See Figure 5. There were no significant main effects of quality of interaction for mentor's in the program on psychological well-being, $F(1, 15) = 0.492$, $p = 0.494$, $\eta 2 = 0.032$. See Figure 6. There were no significant main effects of quality of interaction for mentee's in the program on sense of belonging, $F(1, 19) = 0.638$, $p = 0.434$, $\eta 2 = 0.032$. See Figure 7. There were no significant main effects of quality of interaction for mentor's in the program on psychological well-being, $F(1, 15) = 0.530$, $p = 0.478$, $\eta 2 = 0.034$. See Figure 8. There was a non-significant main effect of the proportion of digital interaction on psychological well-being, $F4, 33 = 0.123$, $p = 0.973$, $n2 = 0.015$. See Figure 9. There was a non- significant main effect of the proportion of digital interaction on sense of belonging, $F4, 33 = 0.543$, $p = .705$, $n2 = 0.062$. See Figure 10.

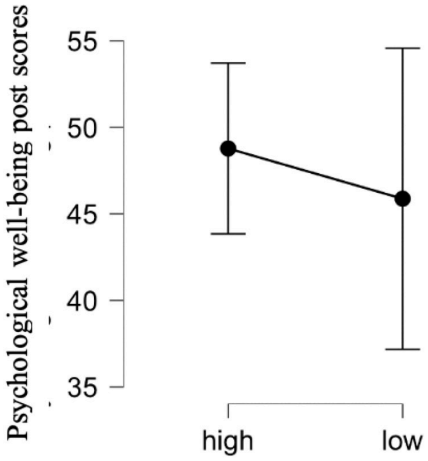


Figure 5: Quality of interaction for mentors in the program & psychological well-being

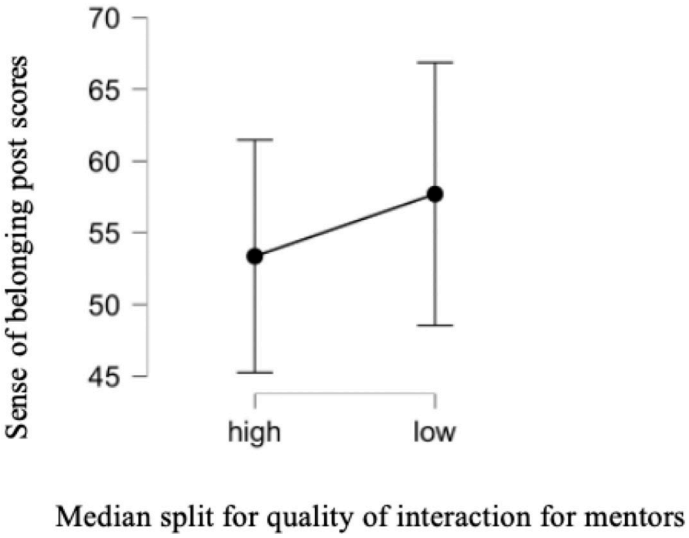


Figure 6: Quality of interaction for mentees in the program & psychological well-being

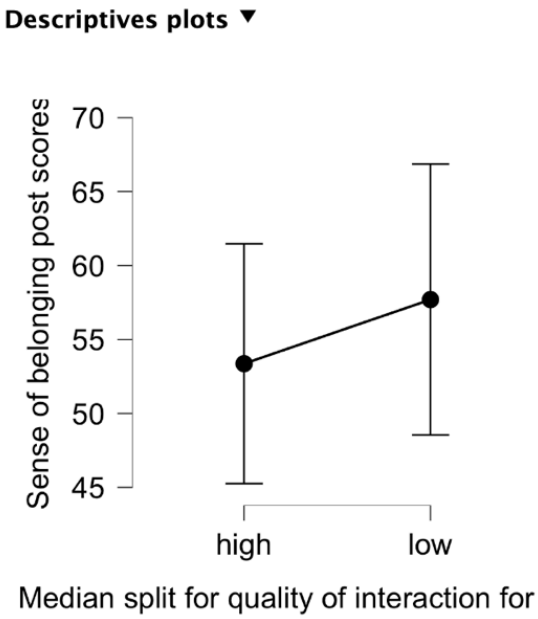


Figure 7: Quality of interaction for mentors & sense of belonging

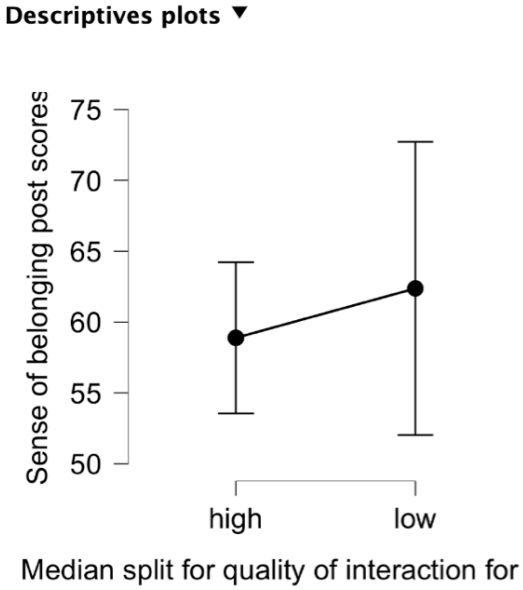


Figure 8: Quality of interaction for mentees & sense of belonging

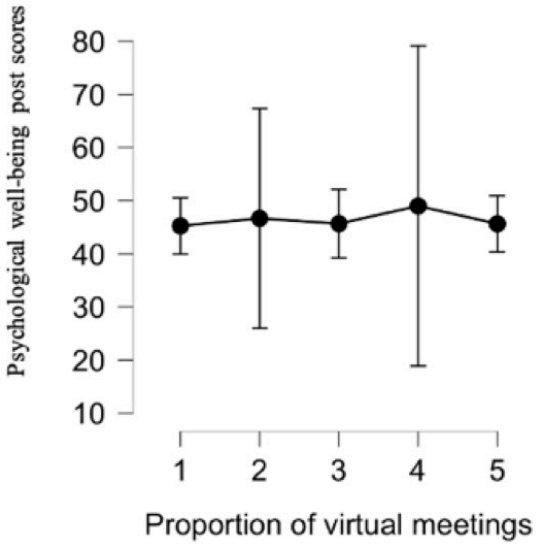


Figure 9: Proportion of virtual meetings & psychological well-being

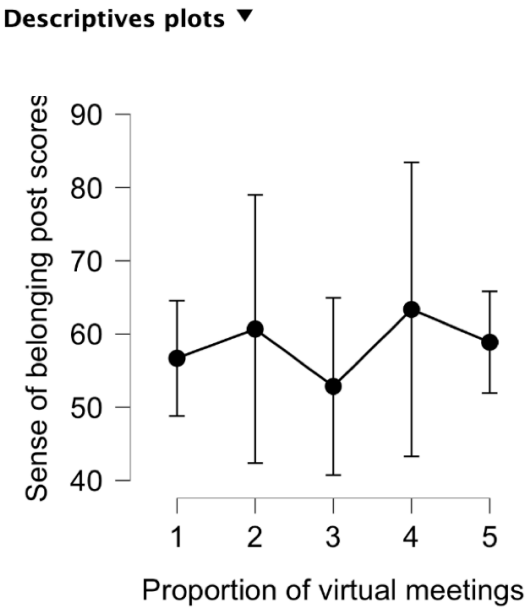


Figure 10: Proportion of virtual meetings & Sense of belonging

Although none of the null hypotheses were disproved, other significant differences were found during analysis. There was a significant effect of participation in the BDB program on psychological well-being scores, $F(1, 54) = 3.116$, $p = 0.083$, $\eta 2 = 0.043$. Importantly, there was a significant interaction effect of participation and time on psychological well-being, $F(1, 54) = 6.852$, $p = 0.011$, $\eta 2 = 0.024$. There was additionally a significant simple main effect of participation on the pre-program scores for psychological well-being, but not the post-program scores, such that participants had much higher scores ($M = 47.553$, $SD = 8.320$, pbonf = 0.007) than non-participants ($M = 41.278$, $SD = 6.632$, pbonf = 0.007). See Figure 11.

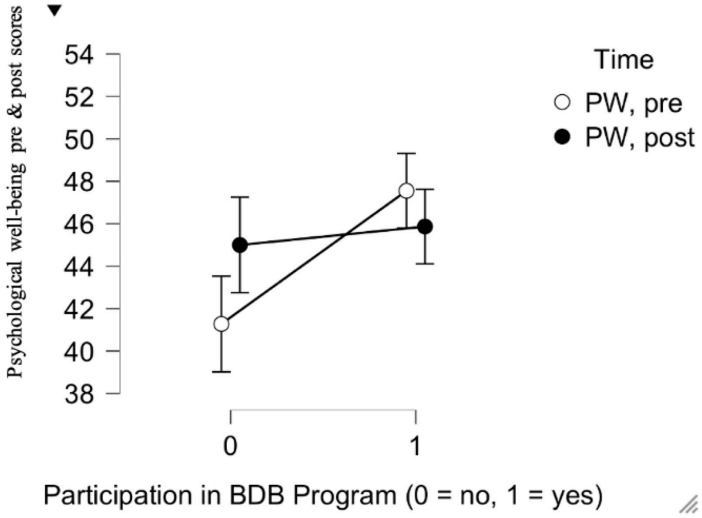


Figure 11: Interaction between participation in the program (0 = no & 1 = yes) & psychological well-being pre- & post-scores

There was also a significant effect of participation in the BDB program on sense of belonging scores, $F(1, 54) = 8.888$, $p = 0.004$, $\eta 2 = 0.091$. Notably, there was a significant interaction effect of participation and time on sense of belonging, $F(1, 54) = 13.136$, $p < 0.001$, $\eta 2 = 0.069$. A significant simple main effect was found for BDB participation on the pre-program sense of belonging scores, but not the post-program scores, such that participants had much higher pre-program scores ($M = 63.763$, $SD = 12.341$, pbonf < 0.001) than non-participants ($M = 48.389$, $SD = 13.111$, pbonf < 0.001). See Figure 12.

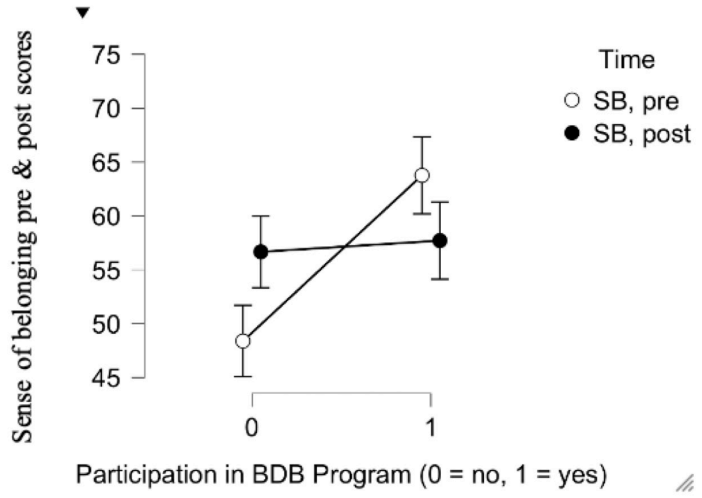


Figure 12: Interaction between participation in the program (0 = no & 1 = yes) & sense of belonging pre- & post-scores

Whether participation began in the Fall 2020 semester or the Fall 2021 semester did have a significant main effect on psychological well-being scores, $F(1, 36) = 4.153$, $p = 0.049$, $\eta 2 = 0.081$. However, there was not a significant interaction effect of semester and time on psychological well-being, $F(1, 36) = 2.263$, $p = 0.141$, $\eta 2 = 0.012$. See Figure 13.

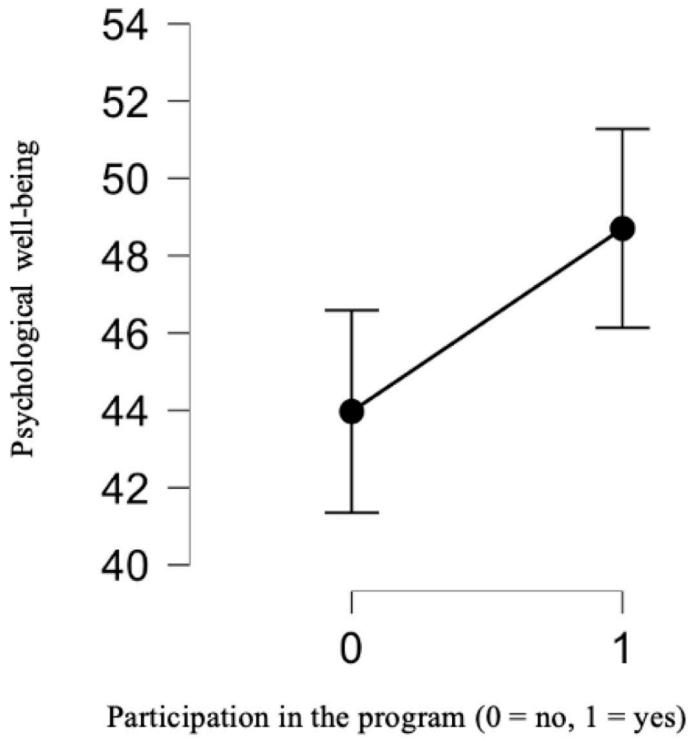


Figure 13: Main effect between participation in the program (0 = no & 1 = yes) & psychological well-being

There were no significant effects of semester on sense of belonging, $F(1, 36) = 0.811$, $p = 0.374$, $\eta 2 = 0.012$. However, there was a significant interaction effect of semester and time on sense of belonging, $F(1, 36) = 6.435$, $p = 0.016$, $\eta 2 = 0.059$. A significant simple main effect was found for semester on the post-program sense of belonging scores, but not the pre-program scores, such that participants of the Fall 2021 school year had much higher scores ($M = 61.364$, $SD = 9.235$, pbonf = 0.018) than participants of the Fall 2020 school year ($M = 52.688$, $SD = 12.370$, pbonf = 0.018). See Figure 14.

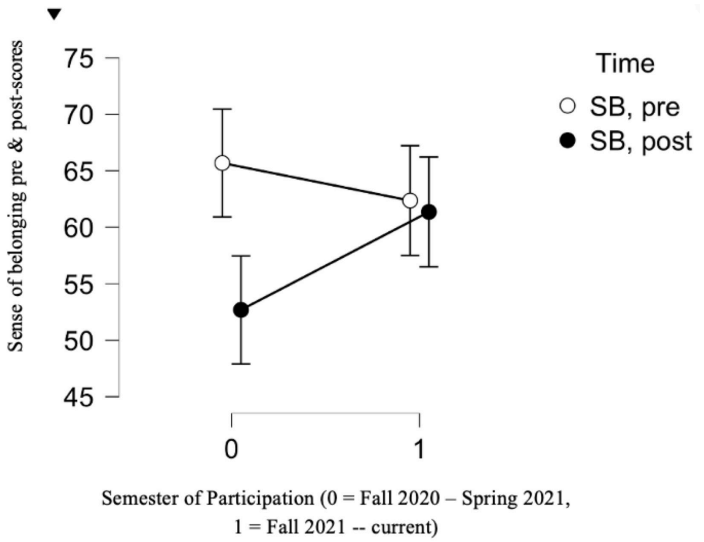


Figure 14: Interaction between semester of participation (0 = Fall 2020-Spring 2021, 1 = Fall 2021-current) & sense of belonging pre- & post-scores

Discussion

The present study examined the efficacy of the Blue Devil Buddies program by measuring participants' scores on psychological well-being and sense of belonging scales. While none of the initial hypotheses proved to be significant, additional analyses revealed significant effects.

The analyses for Hypothesis 1 demonstrated that there was no significant difference in post scores of psychological well-being and sense of belonging between non-participants and BDB participants (mentor or mentee). These results suggest that program participation did not improve these factors (psychological and social well-being) in a significant manner as compared to the general Duke population.
The analyses for Hypothesis 2 showed no significant difference in post scores on the sense of belonging scale between mentees or mentors. This implies that the program did not significantly change mentees' sense of belonging coming into Duke as compared to their respective mentors.

With regards to Hypothesis 3, there was no significant difference in quality of interaction between mentors and mentees and their scores on both the sense of belonging scale and psychological well-being scale. This implies that regardless of their rating of their interaction, their sense of belonging or psychological

well-being did not change.

The analyses for Hypothesis 4 demonstrated that digital interaction did not have a significant effect on participants' scores on both the psychological well-being and sense of belonging scales. Regardless of the degree of their digital interaction, there were no significant differences in their scores for psychological well-being and sense of belonging.

Hypothesis 5 demonstrated that there is not an interaction between perceived quality of interaction and proportion of digital interaction on both scales. This implies that regardless of the amount of digital interaction these mentor/mentee groups had, their quality of interaction was the same.

Although these hypotheses did not yield significant results, additional analysis provided the groundwork for future research. As Figure 12 demonstrates, participation in the BDB program had a significant interaction effect with time on sense of belonging within pre-program scores only. Similarly, participation in the BDB program had a significant interaction effect with time on psychological well-being within pre-program scores only. A possible underlying mechanism for these results might be that the BDB program attracted students who already score high on social and psychological well-being measures. By the end of the year, both participants and non-participants end up with comparable sense of belonging scores and psychological well- being scores, balancing out their initial difference. Ironically, non-participants experience a significant increase in sense of belonging, whereas participants either experienced a zero or decreasing effect on their sense of belonging. One potential explanation is that students naturally adjust to their environment with the passage of time. To be the most effective, programs like BDB must target students with low initial scores on social and psychological wellbeing measures. Further research is necessary to evaluate whether BDB identifies and supports the students who need peer mentorship the most. This study provides the preliminary results to guide future studies.

Semester of participation did not have a significant effect on pre-program scores for sense of belonging, but it did have a significant effect on post-program scores. In the Fall 2020 school year, sense of belonging scores dropped significantly, whereas in the Fall 2021 school year, sense of belonging scores did not change significantly. Further research is necessary to determine the meaning of this. Perhaps the BDB program had protective effects against the added detriment to social and psychological wellbeing caused by COVID-19. Perhaps the program did not have any effect at all, and wellbeing improved naturally as restrictions loosened. To explore this question, future studies must include a greater number of non-participants to confirm the existence and unpack the mechanisms behind possible protective effects.

Semester of participation did not have any significant interaction effects with time, but it did have a significant main effect on psychological well-being. This might mean that psychological wellbeing was generally higher during the Fall 2021 year than the Fall 2020 year among BDB participants, likely due to the return of in-person activities; but further research is required to make any definitive conclusions.

The present study has several strengths, including measuring psychological and social well-being before and after the mentorship program, comparing survey data between mentors and mentees, and comparing survey data between program participants and non-participants. By having pre- and post-program data as well as participant and non-participant data, we were able to conduct a wide array of analyses that other similar studies have not yet done.

In addition to these strengths, it is important to consider the limitations of this study. First, since the study was conducted after the BDB program had been running for several semesters, the survey collected data retrospectively. Retrospective data can be biased due to errors in participant recall of events from one to two years. To address this, one possible recommendation for the BDB program is to provide the Qualtrics survey as part of their onboarding process as well as checking in with their members throughout the semester, compiling real-time data. In this case, participants would only be reporting information based on their current knowledge, feelings, thoughts, and emotions. A second significant limitation is the small sample size of this study. There are a relatively small number of participants in the BDB program to begin. Additionally, only 56 responses in total fulfilled the inclusion criteria. This low response rate meant the analysis had lower statistical power, contributing to the lack of significant

findings. The sample size of future research could be increased by encouraging or requiring BDB program participants to fill out the survey. Although DSG and the BDB program may already be conducting similar surveys, it is important for a third party to collect and interpret results as there could be potential bias in DSG and BDB program evaluations. Future research would also benefit from the growth of the program, as a larger pool of participants would allow for a greater statistical sample size.

This study also utilized a convenience sampling method, which could have introduced confounding variables. Specifically, there is a chance that the people who received the survey who decided to complete it were fundamentally different from other people on Duke's campus who did not receive the study or chose not to respond. As stated above, including a survey as part of the BDB program has the potential to reduce some of these sampling errors. Similarly, the comparison group of non-participant Duke students could be biased. Perhaps a more representative non-participant sample could have been achieved through use of Duke listservs and DukeList.

The last limitation of this study is potential confounding effects of the coronavirus. Due to the fact that the height of data collection was conducted during peak COVID times, it is possible that COVID could have negatively impacted the responses of students. In order to address this, the study could be repeated at a time where the coronavirus has a less significant effect. If the results of these proposed further analyses produce similar results to this study, this can provide greater validity of the results of this study.

This study functions primarily as a preliminary examination of the BDB program. The results have identified several potential correlations and mechanisms that contribute to the efficacy of the program. As future analyses explore these important questions, definitive conclusions about the strengths and weaknesses of the program will hopefully emerge. Further research plays an instrumental role in supporting the social and psychological wellbeing of undergraduate students by illuminating the ways in which the BDB program can improve.

References

1. Asgari, S., & Carter, F. (2016). Peer mentors can improve academic performance: A quasi-experimental study of peer mentorship in introductory courses. *Teaching of Psychology, 43*(2), 131–135. https://doi.org/10.1177/0098628316636288
2. Budny, D., Paul, C., & Bon, L. (2006). The impact peer mentoring can have on freshman students. Proceedings. *Frontiers in Education*. 36th Annual Conference. https://doi.org/10.1109/fie.2006.322596
3. Hoffman, M., Richmond, J., Morrow, J., & Salomone, K. (2002). Investigating "Sense of Belonging" in First-Year College Students. *Journal of College Student Retention: Research, Theory & Practice, 4(*3), 227–256. https://doi.org/10.2190/DRYC-CXQ9- JQ8V-HT4V
4. Holt, L. J., & Fifer, J. E. (2016). Peer mentor characteristics that predict supportive relationships with first-year students: Implications for peer mentor programming and first-year student retention. *Journal of College Student Retention: Research, Theory & Practice, 20*(1), 67– 91. https://doi.org/10.1177/1521025116650685
5. Morrow, J., Hoffman, M., Richmond, J., & Salomone, K. (2002). Revised Sense of Belonging Scale from Hoffman, M. B., Richmond, J. R., Morrow, J. A., & Salomone, K. Investigating "sense of belonging" in first-year college students. *Journal of College Student Retention, 4*(3), 227-256.
6. Pedrelli, P., Nyer, M., Yeung, A., Zulauf, C., & Wilens, T. (2015). College Students: Mental Health Problems and Treatment Considerations. *Academic psychiatry : the journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry, 39*(5), 503–511. https://doi.org/10.1007/s40596-014-0205-9
7. Rhodes, J. E., Schwartz, S. E., Willis, M. M., & Wu, M. B. (2016). Validating a mentoring relationship quality scale. *Youth & Society, 49*(4), 415–437. https://doi.org/10.1177/0044118x14531604
8. Shalaby, R., & Agyapong, V. (2020). Peer Support in Mental Health: Literature Review. *JMIR mental health, 7*(6), e15572. https://doi.org/10.2196/15572

9. Tennant, R., Hiller, L., Fishwick, R. et al. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes* 5, 63 . https://doi.org/10.1186/1477-7525-5-63

Tables

Descriptives table 1

Descriptive statistics for quality of interaction aggregate scores for mentors

	Quality of Interaction Aggregate Score for mentors
Valid	17
Missing	850
Mean	47.588
Median	50.000
Std. Deviation	7.425
Range	23.000
Minimum	36.000
Maximum	59.000

Descriptives table 2

Descriptive statistics for quality of interaction aggregate scores for mentees

	Quality of Interaction Aggregate Score for mentees
Valid	21
Missing	846
Mean	35.905
Median	36.000
Std. Deviation	6.935
Minimum	24.000
Maximum	49.000

Descriptives table 3

Descriptive statistics for psychological well-being post scores

	Psychological well-being post scores
Valid	56
Missing	0
Mean	45.589
Median	44.500
Std. Deviation	7.986
Range	32.000
Minimum	32.000
Maximum	64.000

Descriptives table 4

Descriptive statistics for sense of belonging post scores

	Sense of belonging post scores
Valid	56
Missing	0
Mean	57.375
Median	56.500
Std. Deviation	10.965
Range	47.000
Minimum	30.000
Maximum	77.000

Customizable crying and connection: personalization and its effect on player emotion and empathy

Devinne Moses

Volume 1, Issue 1

Customizable crying and connection: personalization and its effect on player emotion and empathy

Devinne Moses

Office of University Scholars and Fellows, Duke University
https://doi.org/10.55894/dv1.14

Abstract

Player choice sets video games apart from other visual media, and narrative designers work with game developers to implement tools that allow players to make choices and feel involved. Prior research focuses on games that include emotional themes and character customization, which are seen as effective ways to create empathy between player and protagonist. However, most research compares different games with various levels of customization, opening doors to several confounding variables. To address this gap, it is essential to explore how character customization and narrative context influence player emotion, character connection, morality, and engagement in a single game. This study involved participants playing a custom-built interactive fiction game demo (MOSAIC). Overall, players' attitudes toward the demo were consistent across all three experimental groups and the control group, although character connection was higher in groups where participants were given additional story information and/or opportunities to customize the protagonist. This is a valuable step towards understanding video games and player emotion, especially considering gaming's global presence and constant expansion.

Keywords: emotions, characters, morality, video games, customization, self-report

Introduction

Player choice in visual media is often only attributed to video games, and this feature has been a significant staple of this interactive artform over the last few decades. While digital games involve extensive design and development for art and programming, narrative designers are tasked with the overall structure of the story, including options that the player can choose from to alter the game's story beats and endings. In tandem with these decisions, narrative designers are also responsible for constructing stories that highlight engaging and emotional themes of human life. The balance between multiple pathways in games and emotional experiences is difficult to achieve, so understanding how people are affected when they feel involved in a story and its characters is vital for the continuation and expansion of popular, narrative-driven games.

Effective player choice is appealing in most story-based games, especially considering the narrative lore and customization options often found in modern-day games. Disco Elysium (ZA/UM, 2019) is a popular narrative game packed with lore and details only accessible when the player chooses a certain dialogue option, which then creates an emotional experience different from other players. Emotion is a core feature for many narrative-driven games, and hugely successful projects like Life is Strange (Dontnod Entertainment, 2015) combine choice and emotional story beats to craft impactful experiences where the player is involved in constructing the story. Narrative games like these include several factors that contribute to the overall narrative and interactive experience, and this study looks at the choice and customization tools many games offer and asks if they create more heightened, emotional experiences. This study utilizes a self-designed hypertext fiction game, MOSAIC, to test the research question and provide a narrative prototype for participants to play in the study. In this paper, the relevant background literature is explored first, then MOSAIC is briefly described, as well as the

choices the player will make in the game's demo scene. Afterwards, the methods and results from the study is discussed. Finally, the key findings and implications of this study are highlighted before discussing limitations and future research directions.

Background Literature

Emotion and Text-based Media

There has been empirical research on the relationship between emotion and text-based media. Research suggests that readers see fictional characters as real with real human psychologies, and authors can create aesthetic and narrative emotions to enhance realism, attachment to characters, and attachment to the narrative world (Mar et al., 2011). Interactive fiction has increased user control and has looser narrative structure than traditional literature. Research suggests that readers of these texts are transported into a narrative world and act in – rather than witness – the story, encouraging readers to internalize the story through personalization (Green & Jenkins, 2014). Personalization is found most often in the form of character customization, where players get to choose how their character looks, what they wear, and how they behave – such as in games like Animal Crossing: New Horizons (Nintendo, 2020). In fact, several studies indicate that the enjoyment of digital games is predicted by identification with avatars and user-avatar similarity (Elson et al., 2014; Mallon, 2014). Research also suggests that players want to be recognized as an integral part of the game by including choices that lead to branching pathways based on user input, which is a feature popular in many digital visual novel games like 80 Days (Inkle, 2015).

Moral Dilemmas

Previous research on choices in video games has primarily focused mostly on moral dilemmas in video games where players must choose from two or more consequence-heavy options. Findings show that most players, when given a choice, opted for the good or morally right choice in the video game (Consalvo et al., 2019; Tancred et al., 2018; Formosa et al., 2021). Several frameworks have been developed to provide creators with tools to test moral dilemmas and interactivity, and it has been found that most people have trouble making the leap into enjoying being morally evil in a game. Identification with characters in morally challenging stories is made possible through choice tools that developers can implement in their games, such as positive feedback loops or false/deceptive choices (Mawhorter et al., 2014; Chew & Mitchell, 2020; Cole & Gillies, 2021).

Gaps in the Literature

Most games are either crafted with several choices in mind or strictly designed as a linear story, but there is rarely an option to choose one or the other for a single game. This leads to research that compares games in a similar genre that offer either multiple choices or linear experiences to observe player emotion and experience. However, comparing similar games is not the same as comparing the same game, especially considering the potential confounds the games' differences could create. Games differ in art style, story, genre, mechanics, and much more, and variation in these aspects can affect player emotion alongside a game's customization features. This study minimized potential confounding variables by focusing on the same game with minor differences in how the player receives information and makes decisions. This created a more definitive analysis on player choice, information, and experience.

Design Summary

Participants were asked to read a short demo scene from the hypertext fiction MOSAIC and make choices for a character that leads to a certain negative outcome. The dependent variables measured include the participant's emotional reaction to the scene's outcome and how connected the participants felt to the character they played. Participants self-reported emotions and emotional intensity by choosing from a list of 10 discrete emotions (5 positive, 5 negative) taken from the Multidimension Emotion Questionnaire (Klonsky et. al, 2019) and indicated on a scale of 1-10 how strong they felt that emotion after reading the demo scene. Participants self-reported character connection through Likert-scaled character

identification questions pertaining to character empathy and sympathy. Limitations to this style of reporting included unverifiable participant honesty, varied introspective ability, and varied interpretation of emotions and rating scale. Experimental groups involved whether participants saw a pretext scene that contextualizes the story before playing the demo, as well as whether participants got to "customize" their character with accessories described in the demo. One experimental group had only the pretext scene, another group had the character customization options, and another group had both. The control group did not have any personalization options (fig.1). The variables for the experimental groups were based on two determinants of character identification: storytelling techniques and character-based factors. The demo's outcome was the same regardless of a participant's decision or group, but the participants were not aware of that. Exploratory dependent variables were also studied, such as morality and participant engagement. These were measured by asking a series of Likert-scaled, identification questions.

Data was collected through links to the online survey, which was the best method to get the most participants possible while adhering to Duke University's COVID-19 guidelines outlined for summer 2021.

It is hypothesized that including character context and characterization increases participant emotional intensity. Data supporting this hypothesis should clearly demonstrate how video games and hypertext fiction establish player connection compared to other media like film and television. If people feel more connected to these stories, this could encourage innovation in the media entertainment industry to incorporate more interactivity, like seen in recent years with Netflix films such as Black Mirror: Bandersnatch (Slade, 2018). This recent experimentation with interactivity may signal the industry's willingness to adopt and change new formats for success, and it may signal the start of a revolutionary shift for films and other entertainment industries. As video games grow in popularity, audience connection and choice could be the key elements for other media to stay relevant.

MOSAIC

MOSAIC is a narrative or interactive novel game in the Worldview Maturation genre that involves selecting from various text-based choices to influence the character relationships and overall story. The game tells the story of Jamie, a struggling poet in modern-day Fenway-Kenmore, Boston, the home of some of the greatest poetic geniuses of all time. Playing as Jamie, the player makes a series of choices that involve Jamie taking prescribed medication, engaging with his/her/their partner Mehri, and meeting new people on a journey of growth. While Jamie is a fully constructed character, a lot of his/her/their personality depends on the player's decisions. There are minimal visual and audio representations for Jamie, allowing players to shape the character through their in-game choices and potentially project themselves into or role-play as Jamie. For this experiment, participants were informed that this was a demo for a larger game. Participants read a 3-minute scene, and personalization groups were allowed to choose clothing options and music preferences that appeared later in the demo, as shown in Figure 2. Every dialogue choice the player made would lead to a different dialogue response, so players would get immediate feedback based on their choices. However, every choice made still resulted in the same ending for the demo scene, which was unknown to the players. This combination allows players to feel in control of developing Jamie's personality and decision-making without greatly altering the fundamental narrative plot. All participants were given the option of different pronouns that Jamie would identify with (she/her, he/him, they/them), which was implemented to establish a baseline connection between the player and Jamie. See Figure 3 for a sample page from the demo.

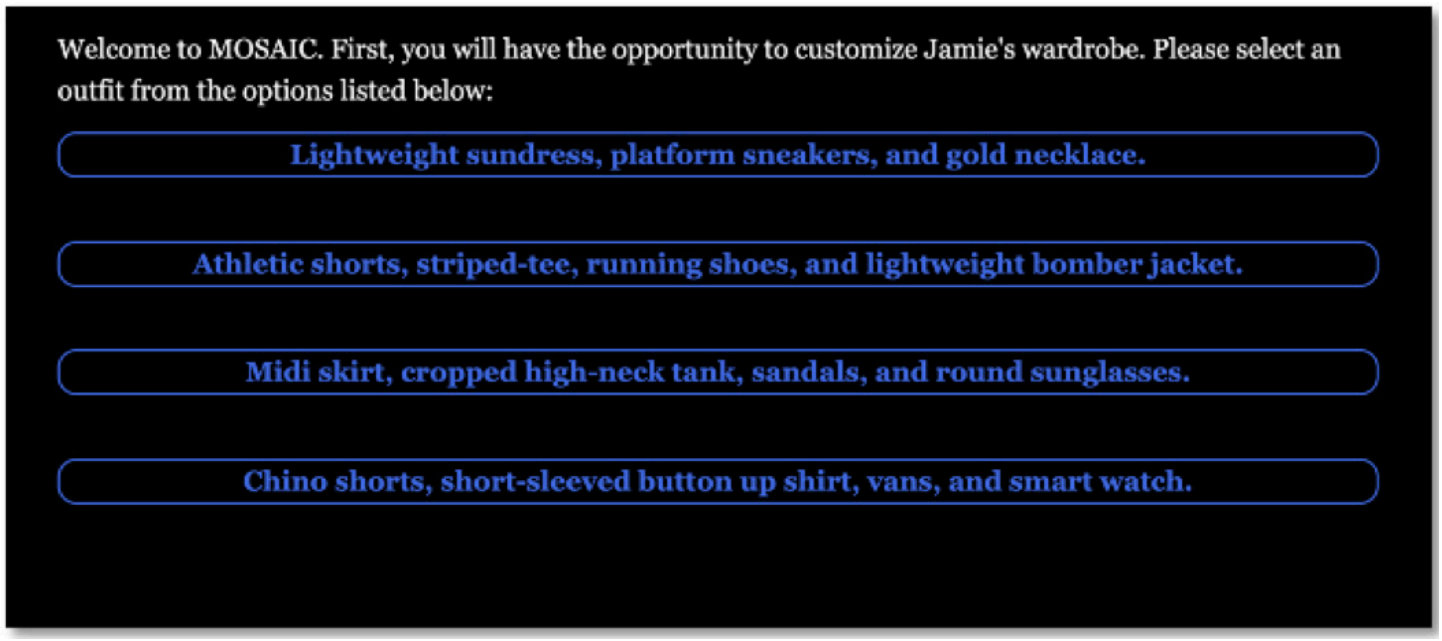


Figure 2: A Page Prior to MOSAIC Scene

Note: These options were reflected later in the story, to show the players' choices were influential.

Figure 3: Page from MOSAIC Demo

Note: Players saw blue outlined words that lead to small interjections on the side of the story.

Methods

Convenience sampling was used to reach undergraduate students, faculty, family, and friends by sharing the survey via social media, email, text, and group chats. To be included in the study, participants had to be over the age of 18. The target sample size was 80 to have approximately 20 participants in each study group, and 77 participants was the actual sample size. Participants were compensated with a $5 Amazon gift card for completing the survey entirely.

The 8 to 10-minute survey consisted of an emotion/intensity table with 10 discrete emotions that asked participants to indicate at least 3 emotions they were feeling and the intensity of those emotions after completing the short scene from the interactive fiction. See Figure 4 to see how the table was formatted for the participants. The survey then includes 5 questions that asked participants how much they connect with the character they controlled, 3 questions about participant's sense of morality, and 3 questions about participant's engagement with the story – all using a Likert scale (5 options ranging from strongly disagree to strongly agree). See Figure 5 for an example of what the questions looked like. Questions were created based on the literature and from established measures like the Multidimension Emotion Questionnaire (MEQ) (Klonsky et. al, 2019).

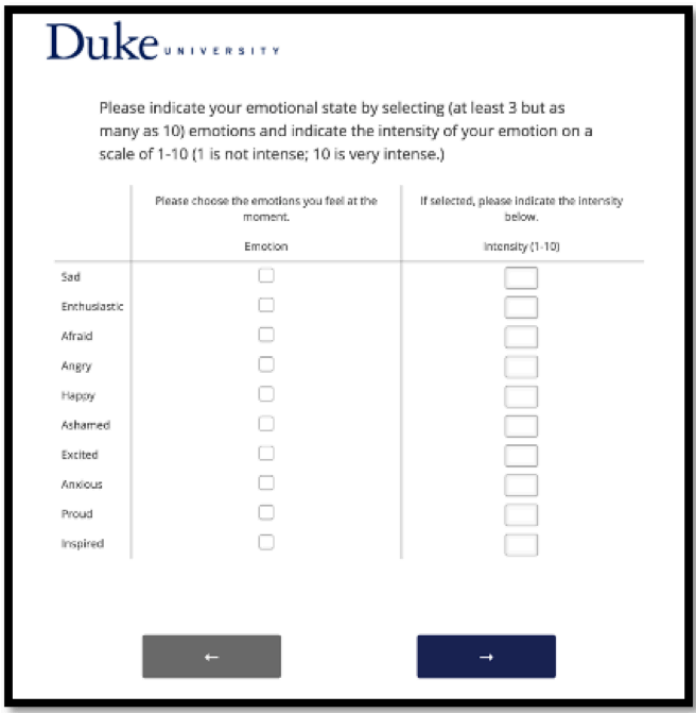


Figure 4: Emotion and Emotional Intensity Response Table

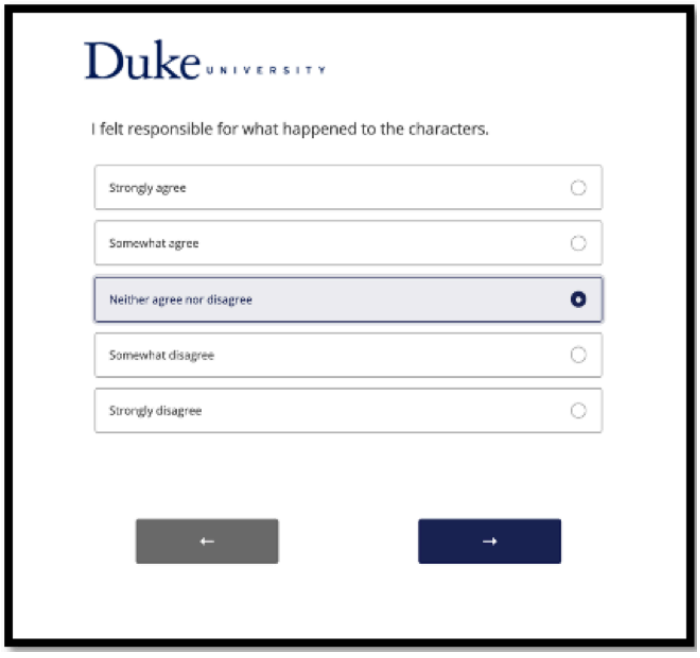


Figure 5: Question from Participant Survey

Four different versions of the study were created so that each survey group represented different combinations of the independent variables. The control group had no pretext before reading the short scene, so they were not given any context pertaining to the story and the character's current situations. The control group also did not have any personalization options, so they were not given the opportunity to pick an outfit for their character or a certain type of music (which would be represented in the story). One survey group had only the pretext, another survey group had only the personalization, and the last survey group had both the pretext and personalization options. All groups received and answered the same questions after reading the short scene.

The emotion questions were taken from the MEQ, which was intended to establish a comprehensive and more accurate measurement for human reports on emotional state. While the MEQ includes variables such as persistence when measuring emotion, intensity was the most applicable to the study since participants reacted to a short scene from a story. The MEQ has been used mainly to assess emotional state in general; however, its flexibility means that it can be applied to the current study.

The character connection questions were derived from readings in the literature that focused on character relatability and decisions in video games (Mar et. al, 2011; Mallon & Lynch, 2014; Elson et. al, 2014; Green & Jenkins, 2014). Questions included how the participants saw themselves in the character they were playing. These questions encouraged participants to think of their relation to the character, which may have change based on what happened in the story. These were developed because there are not many question sets that aim to understand this relationship to the extent of the current study.

The morality questions were also derived from readings in the literature that focused on player morals, morality meters, and making "good" decision in video games (Smethursts & Craps, 2015; Formosa et. al, 2021; Consalvo et. al 2019; Tancred et. al, 2018). These questions encouraged players to reflect on their decisions and their sense of responsibility. Reverse scored questions that discuss regretful decisions were included to prevent biases. These were developed for this study because there was not a complete set of

morality questions that were applicable to the short scene's content.

The engagement questions are closely related to the morality and character connection questions derived from the readings. These questions are categorized on their own because they do not specifically relate to morality or character connection. These general questions about engagement include how willing a participant is to go back into the short scene and change their answer, how much they think they impacted the game, and simply how engaged they felt. Engagement for players can change depending on character and story, so these questions may reveal the effects of the independent variables. These questions were developed for this study to specifically ask about the player's engagement with the short scene.

The study design most closely resembled a posttest-only with control design, where participants were randomized. The intervention was whether participants read the pretext and were given personalization choices, and the observation was how high they scored on the dependent variables. At the beginning of the survey, participants were asked to give informed consent and confirm that they were above the age of 18. Participants agreed and then were randomly assigned to one of the four possible survey variations per the independent variables. The dependent variables were participants' emotions and emotional intensity and character connection, where emotional intensity was scored from -10-10 (to account for negative and positive emotions) and character connection was scored from 1-5. Our exploratory variables included morality and engagement with the story, which were also scored from 1-5.

Upon completing the survey, participants read a debrief of the study design and purpose, including an explanation of the short scene outcome and how their choices did not have an effect. At the end of the debrief, participants were provided with contact information should they have further questions, and participants had the option to provide their email if they were interested in receiving a $5 Amazon gift-card.

Data Analysis Plan

Responses from the survey were scored through Qualtrics, and there were 4 different scores for emotion, character connection, morality, and engagement, which are all continuous variables. Emotions were scored on a scale of 1-10, but negative emotions were flipped to the negative sign (i.e., sad with an intensity of 6 earned a score of -6). Character connection, morality, and engagement were scored on a scale of 1-5, depending on how strongly the participants agreed with a certain statement. Emotional intensity was averaged by the number of emotions participants indicated in the survey to create an Intensity/Amount value. Participants' data sets include the total value of their input, so character connection ranged from 5-25 while morality and engagement each ranged from 3-15. Some items were reverse scored to increase validity.

Initially, a multivariate analysis of variance (MANOVA) was run in JASP, an open-source statistics software, to compare emotion survey score, character connection survey score, morality survey score, and engagement survey score as dependent variables while the differences in experimental groups represented the fixed factor. The results will include several statistics, such as F and p (both used to measure significance), η^2 (the effect size), and Pillai's trace (test statistic also measuring significance). Four single-test ANOVAS (analysis of variance) were run afterwards to see the individual effects of group differences on the dependent variables. The emphasis for the discussion are the main variables – emotion and character connection – while the other variables are discussed briefly. See Tables 1-4 for the variables used in the analyses. Generally, if a significant p-value is found, post-hoc tests (Bonferroni) would be used to protect against alpha error-rate inflation.

Missing data included surveys that were not entirely filled out. Given the structure of the emotion and intensity questionnaire, some participants left the table blank and moved onto the remaining questions. These participants had no effect on the analysis for this variable, which resulted in a slightly smaller sample size compared to the other dependent variables. The incomplete data likely did not affect the ultimate results in a significant way due to the number of participants gathered in total. Surveys that were largely incomplete have not been incorporated in the data whatsoever.

Results

To recap, it was hypothesized that including character context and characterization increases participant emotional intensity. This meant that all experiment groups (groups 2,3, and 4) should have had more negative emotions, higher emotional intensity, and higher character connection compared to the control group.

The MANOVA test of differences among the four groups was not significant, Pillai's trace = 0.17, $F(3,70) = 1.03$, $p = 0.427$. For the ANOVA tests, effect of survey group on emotional intensity, $F(3,70) = 0.27$, $p = 0.848$, $\eta^2 = .01$; character connection, $F(3,73) = 2.40$, $p = 0.075$, $\eta^2 = .09$; morality, $F(3,73) = 0.41$, $p = 0.748$, $\eta^2 = .02$; and engagement, $F(3,73) = 0.46$, $p = 0.712$, $\eta^2 = .02$, were not significant. Participants' responses were not significantly influenced by personalization or pretext when compared to the control group that received neither. See Figures 6-9.

Discussion

Overall, this study found that, contrary to previous research, pretext and personalization options did not have as much of an influence on emotion and character connection as hypothesized. Pretext and personalization also did not have a large influence on the other exploratory variables like morality and engagement.

There was a nonsignificant effect of survey group on emotional intensity. This result may be indicative of the difficulties involved with playing an interactive fiction demo and survey. Participants may have found it difficult to emotionally engage with the demo if they have not experienced a web-based interactive text before. This finding goes against previous research that found personalization and narrative understanding to have a greater effect on a person's emotional experience (Elson et al., 2014).

There was also a nonsignificant effect of survey group on character connection. Figure 1, however, shows a raw increase in character connection for groups 2, 3, and 4 compared to the control group. While this is not a significant effect, this trend suggests that participants felt more connected to the protagonist when they were given the option to customize Jamie's appearance and music taste as well as learn more about Jamie's situation. A larger sample might have statistically supported this observation and resulted in a significant effect. This trend would align with previous research that emphasized how character creators made players feel more immersed in the game to make decisions based on how the player would make them in real life (Mallon & Lynch, 2014; Green & Jenkins, 2014).

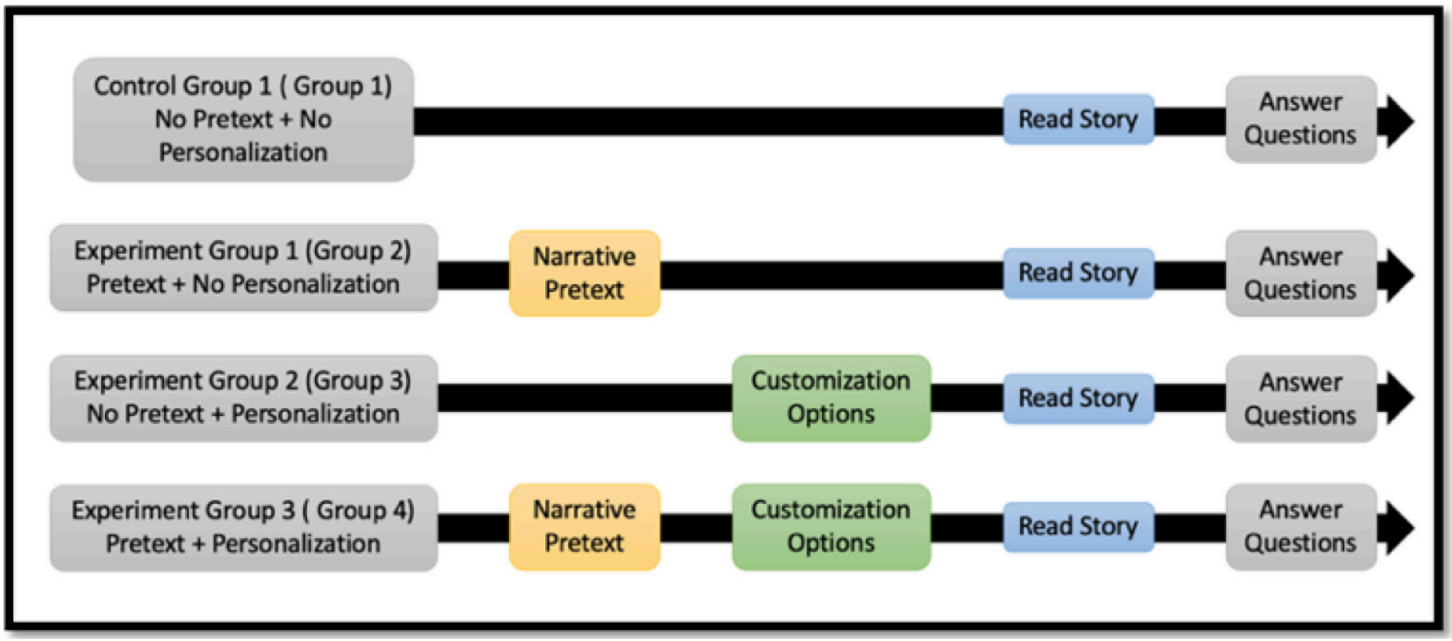


Figure 1: Model Representation of Differences Between Experiment and Control Groups

There was a nonsignificant effect of survey group for the exploratory variables (morality and engagement). For morality, this might be the result because the items implemented were not that morally conflicting compared to other video games, such as Papers Please (3909 LLC, 2013), which at one point has the player decide whether characters can enter a country and reunite with their family despite faulty citizenship papers. Previous research has already concluded that moral challenges in video games can make players regretful or make them always make a good decision (Consalvo et al., 2019). As for engagement, there are a myriad of factors that help video games engage their player base that were not included in the demo, such as visual art or audio. Previous research already concludes that these conventions, among others, enhance player engagement (Chew & Mitchell, 2020).

Conclusion

Strengths & Limitations

The present study was widely accessible given that it was conducted through surveys, which allowed the survey to be shared on various social media. Another strength of this study was the added incentive in completing the survey. Participants received a $5 amazon gift card by completing the survey, which led to higher quality responses.

One of the limitations while conducting the study was that participants needed to open the MOSAIC demo in a separate web browser on a laptop. When looking through the Qualtrics data, several surveys were stopped at that point and were never completed, so response rate decreased because of the added step. Another limitation included the way some of the questions were constructed, particularly the emotion and emotional intensity table. While the survey asked participants to choose at least three emotions and note their intensity, some participants chose less or chose all. Although some of the undesired variation in responses were mitigated by averaging the intensity by the number of emotions chosen, the table could have been better formatted so that participants' responses were as accurate as possible. Given the COVID-19 restrictions, a survey was also the most ideal way to conduct a study without putting participants at risk. However, this opened the study to response biases, where participants might have answered dishonestly or in a way to look socially acceptable. These biases were somewhat mitigated by ensuring participants that the study was anonymous. The survey also had to be distributed through duke listservs, so most responses came from Duke students, and this convenience sampling made the population not representative of the general video game and hyperfiction population.

Future Directions

The present study would be enhanced by including lab work where participants play a game with the necessary equipment to prevent technical difficulties that could take away from the game's emotional potential. During the in-person study, verbal testimonies by participants could offer more context and discussion material to enhance the statistical data, as previous research has done in the past. The study could also benefit from using a more complete video game or interactive fiction instead of a quick demo. A video game or fully designed interactive fiction could offer a more accurate depiction of emotional sensations in gaming, so creating two versions of a well-designed video game that differs based on customization tools would be an interesting pursuit. If cost is a factor, some games now can be used to simulate a similar experience. Role-playing games like Animal Crossing: New Horizons (Nintendo, 2020) offer several cosmetic features for players to customize with, but a new study could assign a group of players to a particular character with a set style of clothes while another group can freely change their character how they see fit. A future study like this could test customization and its effect on emotion and character connection.

Final Thoughts

Although this study did not result in anything substantial in terms of emotion, character connection, and personalization, it did take important steps toward understanding the impact of player preferences and personalization within video games. Studying personalization in video games should use this format more often, where the study uses one game to test character customization rather than different games.

Using the game demo MOSAIC was a novel way to test participant emotion, but this novelty means participants are not necessarily accustomed to this style of testing. The limitations discussed in this study may have led to these results, but a more methodically sound study may have more significant conclusions. Interactive fiction is rarely discussed in modern popular culture, though it is a unique combination of literature and video games that has found a strong, persevering community on various sites. The video game industry saw tremendous growth in 2020 despite the COVID-19 pandemic, and it looks to continue growing in the next several years with several big companies like Amazon, Google, and Netflix competing to break into the market dominated for years by Microsoft, Sony, and Nintendo.

Understanding how video games impact players, especially on an emotional level, may advance the industry and help it branch into new markets or those occupied by modern-day film, television, and literature. This global market will have the opportunity to share relatable and impactful stories that would not only garner critical and commercial acclaim, but also teach people different perspectives and experiences for a broader, more cultured worldview.

References

1. Chaouli, M. (2005). How interactive can fiction be? *Critical Inquiry, 31*(3), 599-617.
2. Chew, E. C., & Mitchell, A. (2020). Bringing art to life: examining poetic gameplay devices in interactive life stories. *Games and Culture, 15*(8), 874–901.
3. Cohen, J. (2006). Audience identification with media characters. *Psychology of Entertainment,* eds J. Bryant and P. Vorderer (Mahwah, NJ: Lawrence Erlbaum Associates), 183–198.
4. Cole, T., & Gillies, M. (2021). Thinking and doing: challenge, agency, and the eudaimonic experience in video games. *Games and Culture, 16*(2), 187–207.
5. Consalvo, M., Busch, T., & Jong, C. (2019). Playing a better me: how players rehearse their ethos via moral choices. *Games and Culture, 14*(3), 216–235.
6. Cova, F. & Garcia, A. (2015), The puzzle of multiple endings. *The Journal of Aesthetics and Art Criticism,* 73, 105-114.
7. Dontnod Entertainment (2015). *Life is Strange [PC Game]*. Square Enix.
8. Elson, M., Breuer, J., Ivory, J. D., & Quandt, T. (2014). More than stories with buttons: narrative, mechanics, and context as determinants of player experience in digital games. *Journal of Communication, 64*(3), 521–542.
9. Formosa, P., Ryan, M., Howarth, S., Messer, J., & McEwan, M. (2021). Morality meters and their impacts on moral choices in videogames: a qualitative study. *Games and Culture*.
10. Green, M. C. & Jenkins, K. M. (2014). Interactive narratives: processes and outcomes in user-directed stories. *Journal of Communication, 64*(3), 479–500.
11. Inkle, Cape Guy Limited (2014). *80 Days [PC Game]*. Inkle, Profile Books.
12. Mallon, B., & Lynch, R. (2014). Stimulating psychological attachments in narrative games: engaging players with game characters. *Simulation & Gaming, 45*(4–5), 508–527.
13. Mar, R., Oatley, K. Djikic, M., & Mullin J. (2011). Emotion and narrative fiction: interactive influences before, during, and after reading, *Cognition and Emotion, 25*(3), 818-833.
14. Mawhorter P., Zegura C., Gray A., Jhala A., Mateas M., & Wardrip-Fruin N. (2018). Choice poetics by example. *Arts*. *7*(3), 47.
15. Mawhorter, P., Mateas, M., Wardrip-Fruin, N., & Jhala, A. (2014). Towards a theory of choice poetics. *University of California, Santa Cruz*.
16. Nintendo (2020). *Animal Crossing: New Horizons [Nintendo Switch Game]*. Nintendo.
17. Pope, L. & 3909 LLC (2013). *Papers Please [PC Game]*. 3909 LLC.
18. Robson, J. & Meskin, A. (2016), Video games as self-involving interactive fictions. *The Journal of Aesthetics and Art Criticism,* 74, 165-177.
19. Slade D. (2018). *Black Mirror: Bandersnatch [Netflix streaming]*. Netflix.
20. Smethurst, T., & Craps, S. (2015). Playing with trauma: interreactivity, empathy, and complicity in the walking dead video game. *Games and Culture, 10*(3), 269–290.
21. Tancred, N., Vickery, N., Wyeth, P., & Turkay, S. (2018). Player choices, game endings and the

design of moral dilemmas in games. CHI PLAY'18 Extended Abstracts, 627-636.
22. ZA/UM (2019). *Disco Elysium [PC Game]*. ZA/UM.

Acknowledgements
I would like to thank the Office of University Fellows and Scholars for funding this experiment. I would like to thank Beatriz Velazquez and Dr. Greg Samanez-Larkin for their guidance and willingness to review my research.

Figures and Table Captions

Figure 1

Model Representation of Differences Between Experiment and Control Groups

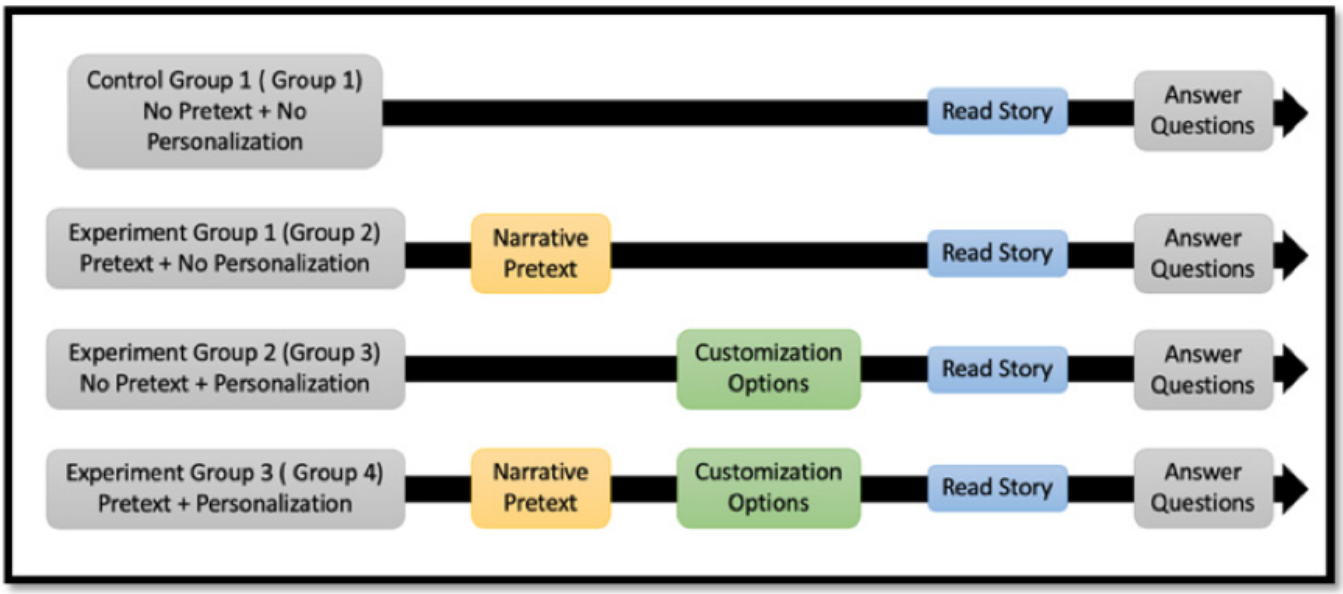
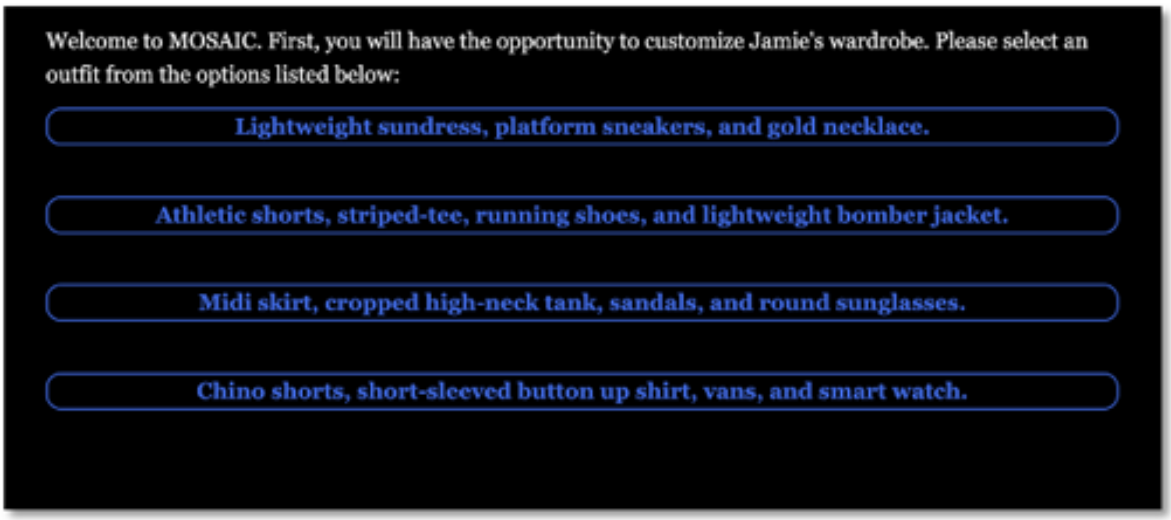


Figure 2

A Page Prior to MOSAIC Scene.



Note: These options were reflected later in the story, to show that the players' choices were influential.

Figure 3

Page from MOSAIC Demo.

Note: Players saw blue outlined words that lead to small interjections on the side of the story.

Figure 4

Emotion and Emotional Intensity Response Table

Figure 5

Question from Participant Survey

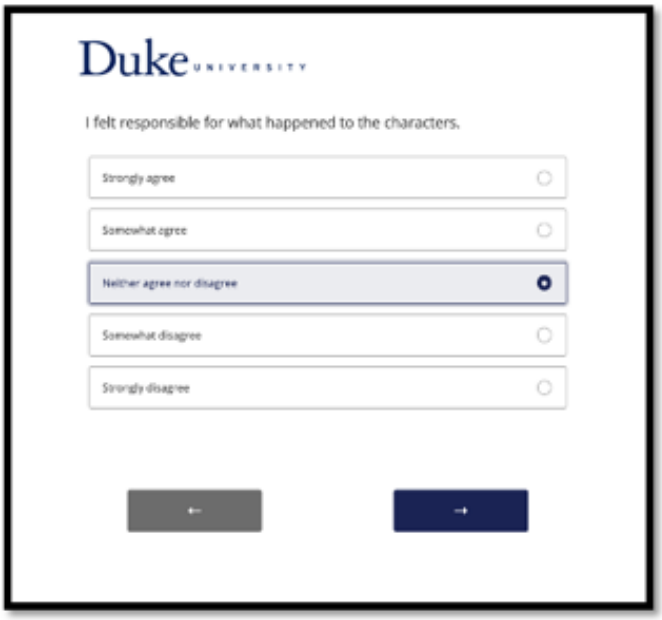


Figure 6

Descriptive Plot Depicting Differences in Character Connection Across Survey Groups

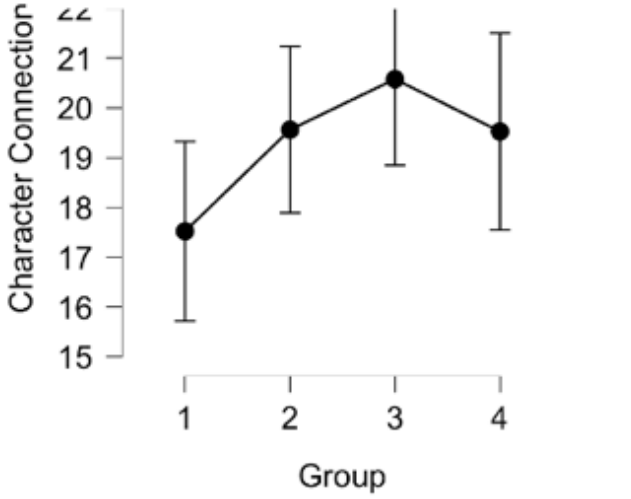


Figure 7

Descriptive Plot Depicting Differences in

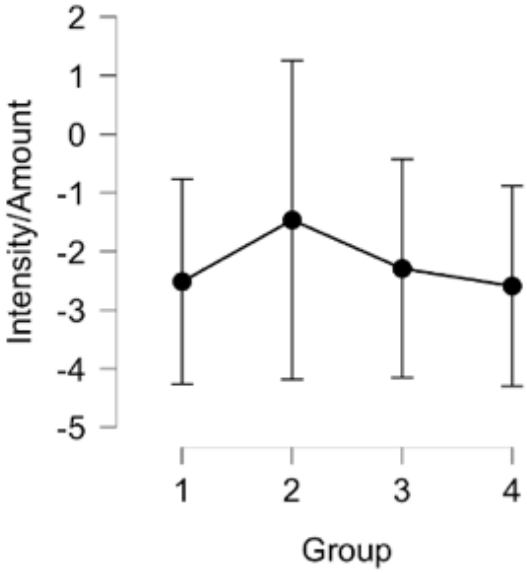


Figure 8

Descriptive Plot Depicting Differences in Morality Across Survey Groups

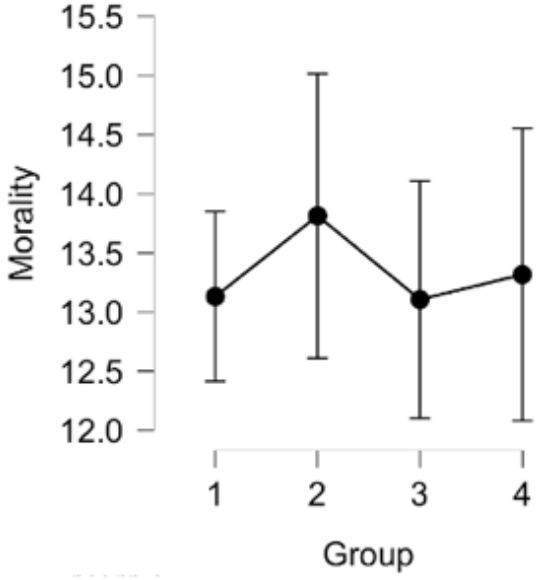


Figure 9

Descriptive Plot Depicting Differences in Engagement Across Survey Groups

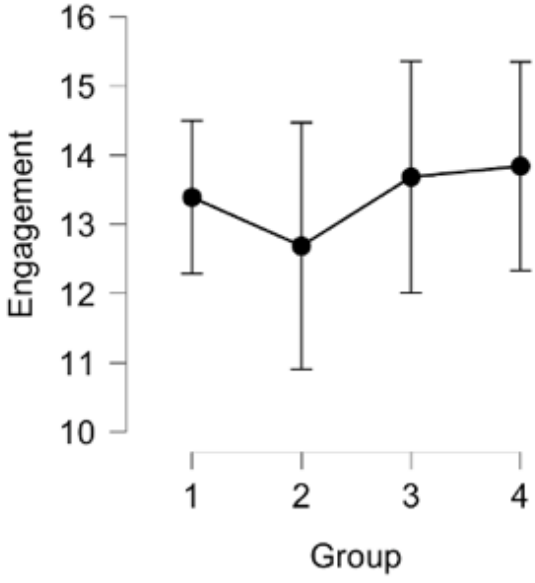


Table 1 & 2

Means and Standard Deviations for Character Connection and Morality Split by Survey Groups

	Character Connection				Morality			
	1	2	3	4	1	2	3	4
Valid	23	16	19	19	23	16	19	19
Missing	0	0	0	0	0	0	0	0
Mean	17.522	19.563	20.579	19.526	13.130	13.813	13.105	13.316
Std. Deviation	4.176	3.140	3.595	4.101	1.660	2.257	2.079	2.562
Minimum	9.000	12.000	12.000	11.000	9.000	9.000	10.000	10.000
Maximum	25.000	27.000	29.000	26.000	15.000	17.000	16.000	18.000

Table 3 & 4

Means and Standard Deviations for Engagement and Emotional Intensity Split by Survey Groups

	Engagement				Intensity/Amount			
	1	2	3	4	1	2	3	4
Valid	23	16	19	19	22	15	18	19
Missing	0	0	0	0	1	1	1	0
Mean	13.391	12.688	13.684	13.842	−2.514	−1.464	−2.292	−2.590
Std. Deviation	2.554	3.341	3.481	3.132	3.944	4.915	3.748	3.547
Minimum	9.000	6.000	4.000	7.000	−8.000	−7.333	−6.667	−8.000
Maximum	18.000	20.000	19.000	19.000	7.000	10.000	6.667	6.250

Acknowledgements

VERTICES EXECUTIVE BOARD

Editor in Chief
Julia Davis

Artistic Director
AJ Kochuba

President
Cydney Livingston

EDITORIAL BOARD

Sasha Bacot
Colby Cheshire
Amishi Desai
Kaeden Hill
Abby Strausbaugh Hjelmstad
Ava Rothrock
Amar Ruthen
Arnav Singh
Alan Wang
Velda Wang
Rachel Washart
Tianrui Wu
Catherine Xiang

DESIGN

Cover Design
AJ Kochuba

Website Design
Sage Cooley

FACULTY REVIEWERS

Alexander Fisher, Ph.D. | Assistant Professor of the Practice of Statistical Science

Fan Li, Ph.D. | Professor of Statistical Science, Professor of Biostatistics & Bioinformatics

Mary K. Rudy, Ph.D. | Professor of Gender, Sexuality and Feminist Studies

Victoria Szabo, Ph.D. | Research Professor of Art, Art History, and Visual Studies

A special thank you to:
Reviewers from Georgetown's Scientific Research Journal