Digital Mental Health: Moderators and Mechanisms of an Online Mental Health Intervention

by

Adam Khan Pettitt

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Psychology

Dissertation Committee:

Dr. Nicholas Allen, Chair

Dr. Alayna Park, Core Member

Dr. Robert Chavez, Core Member

Dr. Beth Stormshak, Institutional Representative

Dr. Valerie Hoffman, External Representative

University of Oregon

Summer 2024

DISSERTATION ABSTRACT

Adam Khan Pettitt

Doctor of Philosophy in Psychology

Title: Digital Mental Health: Moderators and Mechanisms of an Online Mental Health Intervention

In the last few decades, digital approaches to mental health treatment has become more prevalent and widespread in an effort to make mental health treatment more accessible to a wider range of individuals. This dissertation aimed to identify and characterize the moderators and mechanisms of a digital mental health intervention (DMHI). Because of the nascent nature of the field, much of the research that has been conducted has focused on *if* digital mental health interventions are effective. Much of that research has shown it to be on-par with in-person interventions. However, little research has examined the mechanisms by which these interventions are effective.

Across five sub-studies, this dissertation sought to elucidate some of the underlying mechanisms for who this DMHI is effective for, how individuals interact with the DMHI, and identify the underlying mechanisms of improvement in symptoms of depression and anxiety. Participants were drawn from a larger sample of individuals who participated in the Meru Health Program, which is a DMHI platform available to the public. Participants underwent an 8-week or 12-week intervention (depending on which version they were given) that focused on therapeutic techniques derived from Cognitive Behavioral Therapy, Mindfulness Based Stress Reduction, and other evidence-based therapies. Participants were administered demographic questions at the beginning of the intervention and administered depression and anxiety questionnaires at enrolment and every 2 weeks until the end of treatment. The analyses used in this dissertation were mixed-methods ranging from mixed-effects modeling to qualitative thematic analyses aimed at understanding the underlying mechanisms for efficacy within the MHP. Results from Chapter 1 Study 1 revealed that across age, gender, and race, the DMHI was effective for all groups, and in particular (from Chapter 1 Study 2) there was a disproportionate drop in suicidality within gender expansive individuals when compared with cis gender individuals. Additionally, results from Chapter 2 study 1 indicate that participants engaged in messaging with their therapist for a wide array of reasons, including rapport building and solving tech difficulties. Further analyses in Chapter 2 Study 2 revealed that, within the first week, days active within the app was the most predictive of completion of the DMHI. Finally, the results from Chapter 3 demonstrated that improvements in HRV across the DMHI are associated with reductions in depressive symptoms. The implications of these findings and proposed areas for ongoing research are discussed.

Keywords: digital mental health, depression, anxiety, treatment moderators, heart rate variability

ACKNOWLEDGEMENTS

I would like to take this space to acknowledge the people who have truly helped me in ways that I can never really thank them enough for. First, I would like to thank my parents, Sajida and Donald Pettitt. You both have been a steadfast form of support whenever I have needed it and I know that you would be there for me through anything. The security from knowing that I could rely on you in any struggle allowed me to take the risks necessary for completing this degree. I would also like to thank my friends who have supported me through the dissertation writing process. Rosario Peralta, since we became friends over a decade ago, you have been the most incredible friend and I cannot imagine where I would be without your constant friendship. Drew Terhune, I met you at the start of my journey at UO and your insights and friendship have helped shape who I am as a person and as a therapist. Dr . Sara Lieber, your friendship and the laughter we shared during the pandemic (and before) were a panacea to the ups and downs of graduate school, and I'm so grateful for our friendship.

To my committee, Dr. Alayna Park, Dr. Beth Stormshak, Dr. Rober Chavez, Dr. Valerie Hoffman, and Dr. Nick Allen, I would also like to extend a heartfelt thank you for your support and feedback. In particular, I would like to thank Dr. Valerie Hoffman, who gave me a chance at Meru Health that started this entire journey. You were the best boss I have ever had and you let me flourish as a researcher and as a person in a new environment and I will always be so grateful. To Dr. Nick Allen, words truly cannot express how grateful I am for everything you have done for me. Even though you are my advisor, I truly see you as a mentor, a friend, and a confidant. You helped guide me through some of the most tumultuous times of my life and truly I would not have finished this program if it were not for you. I will always be grateful for your guidance and support.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1

General Introduction

Major depressive and anxiety disorders are common mental difficulties associated with significant impairment in psychosocial functioning and reductions in quality of life (Hasin et al., 2018; Olfson et al., 2019). Many mental health interventions have been shown to be effective in addressing these conditions, with clinical and epidemiological studies indicating lower rates of symptom relapse and recurrence compared to usual care and control conditions (Carpenter et al., 2018; Cuijpers et al., 2021; van Dis et al., 2020; Z. Zhang et al., 2018). Unfortunately, it is estimated that more than half of people in need of treatment do not receive care (Walker et al., 2015; Wang et al., 2005). Mental health interventions also remain inaccessible to many populations due to lack of affordability, transportation issues, stigma, and other structural barriers (Chekroud et al., 2018; Kessler et al., 2001). Additionally, some clients can avoid or disengage from in-person treatment because there is a mismatch between what is being offered and what is desired by the client, creating issues with overall engagement with available mental health treatments. Digital mental health interventions (DMHI) have become increasingly available through high-speed internet and the continued uptake of mobile technologies that can facilitate telephone, video, and text-based approaches to care (Torous & Haim, 2018). The purpose of this chapter is to provide background on the landscape of digital mental health and provide a rationale for the aims explored later within this dissertation.

Additionally, with the onset of the COVID-19 global health pandemic, there has been an increased need for Digital Mental Health Interventions (DHMIs) due to both the increased mental health burden that it has caused as well as increased desire for much of our daily life to be conducted within the context of digital spaces. However, because these interventions are so nascent, research is still trying to understand which both who benefits from these interventions and the underlying mechanisms for how these interventions work. Recent research has shown that digital interventions often have similar clinical outcomes to their online counterparts (Novella et al., 2022). However, the benefit of DMHIs is that they can be more accessible to a wide group of people that may not be able to or want to participate in traditional mental health services. For instance, individuals who still experience stigma when seeking out medical services (Benuto et al., 2019) may have embarrassment around going to a physical location which could jeopardize their privacy. Additionally, physical (e.g., rural) and cost barriers (e.g., childcare) prevent large groups of individuals from being able to readily participate, despite having the motivation to do so (Bornheimer et al., 2018; Summers-Gabr, 2020). It is not as though well validated and effective interventions have not been developed. Rather, from a public health perspective, often these interventions are simply not accessible to the groups that may want to seek out services. Further research must be conducted into who DMHIs are effective for and whether these groups are seeking out interventions within a digital context.

Multiple studies have examined the relationship between DMHIs and depressive symptoms. One meta-analysis conducted analyses on DMHI RCTs and found that the interventions significantly reduced depressive symptoms compared to control conditions, suggesting that DMHIs can be a promising self-management tool for depression (Firth et al., 2017). Additionally, Donker et al. (2013) conducted a systematic review examining the efficacy of DMHIs on different mental health symptoms and found similar outcomes. DMHIs were related to significant reductions in depressive symptoms, anxiety symptoms, substance use, and reported stress. These findings indicate that depression and anxiety symptoms can be addressed by easily-accessible DMHIs.

However, just because these DMHIs can be accessed by a wide range of individuals does not mean that they are being utilized by different populations. For example, a recent review found that, while different groups often had the same access to DMHIs, there were disparities in utilization (Schueller et al., 2019). For example, Black and Latinx populations were equally likely to have access to smartphones when compared with White populations, but both populations were less likely to be included in clinical research. Additionally, there are often only English-speaking options for these DMHIs, limiting some of the accessibility, especially within the Latinx community. Additionally, another review found that, while age, race and ethnicity were reported in over 30% of DMHI RCTs (Kirvin-Quamme et al., 2023), many of these only reported the proportion in the population and used these variables as a covariate. Very few studies examined these demographic factors as a predictor of clinical outcomes, and none, to my knowledge, examined between group effects within these demographic factors. Even more concerning was the reporting of gender and sex. Fewer than 5% of studies even reported participant gender and sex demographics that included options outside the binary of "Men" and "Women." Again, none of the studies examined between group differences outside of the gender binary. This dearth of research is appalling, given that DMHIs are thought to be an excellent way of reducing stigma and increasing access; but as of right now there is still scant research related to these groups. More research is vital for understanding if these interventions are effective across these different populations.

Once an individual has started one of these programs, research has shown that getting them to complete the intervention can be difficult (Obikane et al., 2022). Because of this, one of the main criticisms of DMHIs is that they are effective at being widely available to people, but not keeping them engaged once they are participating. This could be because often having someone that they are physically accountable to (e.g., a traditional therapist in an office) provides an incentive for the individual to remain engaged. When individuals are removed from having a working connection with someone, it may prove more difficult to remain enthusiastic about the work.

In terms of completion rates, a systematic review by found that adherence rates varied across interventions, with approximately 50% of participants adhering to the intervention on average. The review also highlighted the influence of design factors on adherence rates. Higher numbers of persuasive technology elements and higher levels of dialogue support were associated with increased adherence to web-based interventions (Karyotaki et al., 2018). However, the effect sizes were small, indicating that other factors may also play a role in adherence rates. Further research can illuminate how individuals are engaging with the content and therapists within the app and how that is related to completion of the DMHIs they are enrolled in.

Finally, while these interventions are effective at reducing mental health symptoms, the underlying mechanisms for these effects remain relatively unexplored. In particular, for the DMHI that will be explored for this dissertation, one of the distinguishing factors is the implementation of Heart Rate Variability Biofeedback (HRVB) exercises, which involves providing participants with a wearable device that provides real-time visualization of beat-to-beat heart rate intervals during slow, paced breathing. Recently, there has been a push to provide HRVB to a wider range of individuals. As such, more cost effective Bluetooth heart rate sensors have entered the market (e.g., KYTO HRV Monitor), making HRVB much more accessible. During HRVB practice, patients are shown real-time heart rate data and prompted to breathe at a prescribed rate which corresponds to "resonance" frequency, which increases HRV or heart rate oscillations (Lehrer, 2013). This resonance frequency is thought to improve physical and emotional resilience by strengthening homeostatic functions through increasing cardiac vagal activity and exercising the baroreflexes. As this is such a defining feature of the DMHI that will be used for this dissertation, determining whether this is one of the underlying mechanisms of clinical change could be an important component of understanding the efficacy of this treatment.

As these interventions become more and more relevant, it is not enough for us to simply know that these interventions are helpful in reducing the burden of mental health. Often, these interventions offer different components and appeal to a different population than traditional in-person therapies. It is important to understand both who these interventions are helping and how they are achieving the goal of reducing mental health symptoms. For this dissertation, we will use data collected from a DMHI (described below) and examine the moderators of treatment engagement and success, as well as looking at underlying mechanisms of clinical improvements. Throughout the course of this dissertation there will be 3 separate high level aims across 5 separate studies.

- **Aim 1**: Examine the sociodemographic factors associated with reductions in mental health symptoms in participants undergoing treatment in a Digital Mental Health Intervention.
- **Aim 2**: Examine an array of different engagement factors that are related to completion rates within a Digital Mental Health Intervention.
- **Aim 3**: Examine a hypothesized underlying biophysical mechanism of change for a Digital Mental Health Intervention.

General Methods

Description of Intervention:

Because all 5 studies draw from the same larger sample, I have decided to include in this section a general methods section that details the shared procedures between all of the studies. The Meru Health Program (MHP) is a therapist-supported 8-week and 12-week Digital Mental Health Intervention (DMHI) with evidence-informed components delivered asynchronously via a smartphone app (28,29), including practices derived from mindfulness-based stress reduction, mindfulness-based cognitive therapy, CBT, and behavioral activation therapy treatment protocols. The content was designed to teach participants skills based on mindfulness meditation and cognitive behavioral therapy. The content for each week unlocked automatically at the beginning of each new week, without the need to view prior content and complete prior exercises. Participants were considered to "complete" a week if they had watched the ~50 minutes worth of videos at the start of each week. Participants who completed at least 50% of the weeks (i.e., watched half of the weekly videos) were considered "completers" of the MHP.

The MHP incorporates a continuous care model that includes frequent interaction (up to multiple times a day on an as needed basis) with a dedicated, licensed clinical therapist and as-needed consultations with medical doctors, including psychiatrists. Mental health clinicians at the MHP are required licensed for the state in which they see participants and the clinicians have a variety of different types of licensure (e.g., Licensed Clinical Social Worker, Licensed Professional Counselor). When therapists are onboarded, they undergo a multi-week training that gives them information on both the platform and the models used within the intervention. The Meru Health therapists provided ongoing individual support as needed and curation of the group chat during the DMHI. The interaction took place primarily via chat messaging and, in a few instances, phone calls.

Participants

Though participants could come through multiple avenues (such as specific study recruitment or Meru employee referrals), the vast majority of clients were referred via either their employer, healthcare providers, or employee assistance programs. Meru's main business strategy was aimed at partnering with insurance companies and employers that would then refer people within their institution to the MHP. Participants provided consent at their intake session where they were told that their data would be collected and used to both improve the business product as well as used in future published research. Inclusion/exclusion criteria of the MHP require patients to have at least mild levels of depression (Patient Health Questionnaire-9 > 4) or anxiety (Generalized Anxiety Disorder-7 > 3), own a smartphone, and not have an active substance use disorder, severe active suicidal ideation with a specific plan, severe self-harm, or a history of bipolar disorder or psychosis, as assessed by the clinician during the intake session. Additional details about the program may be found elsewhere (22,28,29).

Depressive and Anxiety Symptoms. Depressive symptoms were measured at baseline and every 2 weeks through the program's end by the Patient Health Questionnaire-9 (PHQ-9), a widely used instrument used to screen for depression (Kroenke et al., 2001). The PHQ-9 consists of a list of nine depressive symptoms with response options ranging from 0 (not at all) to 3 (nearly every day). The PHQ-9 has excellent internal consistency (Cronbach's α of 0.89 in primary care settings), and test-retest reliability (Arroll et al., 2010).

For participants in the earlier 8-week program, anxiety symptoms were measured at baseline and every 4 weeks through the program's end by the Generalized Anxiety

Questionnaire-7 (GAD-7). For participants in the current 12-week program, anxiety symptoms were measured at baseline and every 2 weeks through the program’s end by the GAD-7, a widely used instrument in outpatient and primary care settings to screen for the presence and severity of an anxiety disorder. The GAD-7 has excellent internal consistency and test-retest reliability (Löwe et al., 2008; Spitzer et al., 2006).

Overview of Studies

Chapter 1 Study 1 examines differences across demographic factors for depression and anxiety scores as well as completion rates. Chapter 1 Study 2 examines how suicidality trajectories differ between cisgender and gender expansive participants. Chapter 2 Study 1 uses a mixed-methods approach to understand the ways in which participants are engaging with the intervention and their therapist within the first week. Chapter 2 Study 2 uses similar data from the first week of the intervention to determine predictors of program completion. Finally, Chapter 3 examines how changes in HRV accompanying HRVB practices are associated with reductions in depression and anxiety symptoms.

Chapter 2

Introduction

In recent years, the landscape of mental health care has undergone a profound transformation with the rapid advancement of digital technologies. DMHIs have emerged as promising tools for addressing the global burden of depression and anxiety. These interventions offer scalable and accessible support, allowing individuals to manage their mental health in the privacy of their own spaces, transcending geographical and temporal barriers. However, despite the growing popularity and potential of these digital interventions, there exists a pressing need to understand the nuanced interplay between sociodemographic factors and their effectiveness.

Depression and anxiety are among the most prevalent mental health disorders worldwide. In terms of specific prevalence rates, Reynolds & Kobak (1995) reported that 1-month prevalence rates for major depression ranged from 2% to 3%, and the point-prevalence rate for major depression was between 5% and 9% for women and between 2% and 3% for men, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV).The significant impact of these conditions on individual well-being, as well as the burden they place on healthcare systems, underscores the urgency to develop effective interventions that cater to diverse populations. While digital mental health interventions hold promise, it is crucial to assess how sociodemographic factors, such as age, gender, race, and ethnicity may influence the accessibility, engagement, and outcomes of these interventions.

For instance, Black and Latino populations tend to experience more severe mental health issues that are more debilitating than other racial groups (Breslau et al., 2005). Additionally, it is estimated that up to 20% of people over the age of 55 experience some sort of mental health issue (American Association for Geriatric Psychiatry, 2004). With issues that come with reduced mobility, increased chronic health problems, and other contextual factors, aging can lead to difficulties in life that can have an impact on an individual's well-being and mental health.

In regard to gender, gender minorities are more likely to have higher levels of depressive and anxiety symptoms than their cis-gender counterparts (Reisner et al., 2016). Multiple studies have demonstrated that gender expansive individuals are likely to report higher thoughts of suicidality than cisgender individuals. One study (Reisner et al., 2015) that transgender youth had a two- to three-fold increased risk of depression, anxiety disorder, suicidal ideation, suicide attempt, self-harm without lethal intent, and both inpatient and outpatient mental health treatment compared to cisgender matched controls. Additionally, Perez-Brumer et al. (2017) conducted a population level study and found that in a population of California youth, transgender youth had a significantly higher prevalence of suicidal ideation compared to non-transgender youth. Additionally, while there are some studies that examine these demographic factors as it relates to reductions in symptoms within the context of DMHIs, vanishingly few studies ever publish their findings related to completion and these sociodemographic factors (Kirvin-Quamme et al., 2023).

Taken together, these findings are alarming in that sociodemographic factors can be indicators of more severe mental health issues (caused by more upstream factors). In addition to these concerning findings, racial minorities and gender minorities are much less likely to be in treatment for a variety of reasons (as discussed earlier). Fortunately, more recent research has shown that DMHIs are beginning to show promise for feasibility and acceptability in increasing access that marginalized individuals have to effective mental health interventions (Schueller et al., 2019). The purpose of this study is to determine if this specific intervention is effective across different sociodemographic factors (i.e., age, gender, and race). More specifically, we are seeking to examine how different individuals enter into the MHP, continue throughout the MHP, and finally how they leave the MHP. While we did not have any specific hypotheses related to these outcomes, given the lack of research currently available, we wanted to explore between group relationships as it relates to reductions in mental health symptoms and completion of the DMHI.

Study 1

- **Aim 1**: Examine between-subjects effects for depression and anxiety scores across all time-points within racial and ethnic groups.
- **Aim 2**: Examine between-subjects effects for depression and anxiety scores across all time-points within gender groups.
- **Aim 3**: Examine between-subjects effects for depression and anxiety scores across all time-points within age groups.
- **Aim 4**: Examine completion rates across race/ethnicity, gender, and age groups.

Study 2

- **Aim 1**: Determine if there is a difference in baseline suicidality comparing gender expansive and cisgender participants**.**
- **Aim 2**: Determine if there are differences in trajectories in suicidality between gender expansive and cisgender participants in a DMHI.

Study 1

Method

For this study, it will be in two parts. For the first part, we will examine the clinical characteristics of different sociodemographic groups across the duration of the program.

Specifically, we will look at race, age (binned into a categorical variable), and gender. We have asked participants to self-report their sociodemographic data and will use that to create categorical variables to assess different baseline characteristics, trajectories, and clinical outcomes. Additionally, we will look at other outcome measures, specifically completion rates, to determine if certain groups are completing at the same rate as others.

Because there is a known suicide risk in gender-expansive individuals, we will also be conducting a second part to this study, targeting suicidal ideation as outcome of interest within this population. In order to assess this, we will isolate the 9th item on the PHQ-9 (a suicide related item) to determine how this group differs in their baseline suicidality, the trajectory of their suicidality, and finally outcomes at the end of the intervention.

Participants

Data was collected from a sample of 4722 individuals who started in the MHP between January 1st, 2021 to January 1st, 2023. Over the course of this time, the way that certain variables have been asked have changed. Due to this, there are 24 separate permutations for how sex and gender could combine. A participant's reported gender took precedent over their reported sex, though if their gender was not reported than their sex was used as their gender. If a participant did not answer or chose "Prefer not to disclose," then they were marked as "Other." Due to a low group size, 18 were removed from the dataset and subsequent analyses if they were marked as "Other". Regarding Race and Ethnicity, participants were given two separate question to ascertain their race or their ethnicity (e.g., Hispanic). Participants were assigned whatever race they reported themselves, unless they indicated that they were "Hispanic," in which case they were assigned "Hispanic," regardless of reported race. 142 individuals who marked "Prefer not to disclose" (81) "American Indian or Alaska Native" (20), or "Decline to state" (41) were removed from the analyses because of the low group size. For age, consistent with federal reporting standards for adults, we used the following age brackets to bin participants in: 18-25, 26-34, 35-44, 45-54, 55-64, 65+.

Statistical Analysis

All statistics were conducted using the statistical program R (R Foundation for Statistical Computing, 2018). For the first part of this study, we will run separate mixed effects models examining each demographic factor (i.e., Race, Age, and Gender) in order to determine if there are differences across these groups for PHQ-9 and GAD-7 scores. For each demographic factor, we will run the 2 model: one as PHQ-9 and one as GAD-7 as the outcome variable. Within these models, Time will be a factor variable (as PHQ-9 and GAD-7 are collected bi-weekly), and each demographic variable will be used as a moderator. Post-hoc comparison analyses will be reported in order to determine the differences between groups and across different timepoints. In order to determine if there are differences in completion rates for these demographic groups, we will run separate logistic regression models with completion as the outcome variable.

Table 1.
Summary of descriptive statistics for the participants used in analyses.

Characteristic	Overall, N = 4,723[1]	Asian, N = 645[1]	Black or African American, N = 245[1]	Hispanic, N = 311[1]	White, N = 3,522[1]	p-value[2]
Age	39 (11)	35 (8)	37 (10)	36 (10)	41 (11)	<0.001
Gender						
Man	1,180 / 4,723 (25%)	169 / 645 (26%)	51 / 245 (21%)	88 / 311 (28%)	872 / 3,522 (25%)	
Non-binary	60 / 4,723 (1.3%)	4 / 645 (0.6%)	1 / 245 (0.4%)	3 / 311 (1.0%)	52 / 3,522 (1.5%)	
Woman	3,483 / 4,723 (74%)	472 / 645 (73%)	193 / 245 (79%)	220 / 311 (71%)	2,598 / 3,522 (74%)	
Completion						0.005
Completer	3,575 / 4,723 (76%)	482 / 645 (75%)	165 / 245 (67%)	227 / 311 (73%)	2,701 / 3,522 (77%)	
Drop-out	1,148 / 4,723 (24%)	163 / 645 (25%)	80 / 245 (33%)	84 / 311 (27%)	821 / 3,522 (23%)	
Baseline PHQ-9	11 (6)	10 (6)	12 (6)	12 (6)	11 (6)	<0.001

[1]Mean (SD); n / N (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Results

Race & Mental Health Symptoms

See Table 1 for descriptive statistics of participants. A linear mixed-effects model was conducted to determine whether there were significant differences in PHQ-9 and GAD-7 across multiple racial/ethnic groups. In order to determine whether there were significant effects for these factors, we ran an ANOVA on the linear mixed-effects models (see Table 2). For the PHQ-9, there was a significant main effect for race $F(3,4{,}866.13) = 7.60, p < .001$, as well as a significant main effect for Time, $F(6,17{,}697.09) = 648.53, p < .001$. This indicates that, independent of time, there was a significant effect of race/ethnicity on PHQ-9. Post-hoc analyses showed that Asian individuals ($Estimated\ Marginal\ Mean\ (EMM) = 9.70, 95\%\ CI[9.25,10.15]$) had a lower PHQ-9 score compared to all other racial/ethnicity groups ($EMM = 10.73, 95\%\ CI[10.45,11.00]$; see Figure 1). There was also a significant main effect for Time, indicating that, as time increased, PHQ-9 scores decreased. In particular there was a significant difference from baseline ($EMM = 9.70, 95\%\ CI[9.26,10.15]$) to end of treatment ($EMM = 3.88, 95\%\ CI[3.34,4.43]$). Additionally, there was a significant interaction effect for time and race/ethnicity, $F(18,17{,}724.67) = 2.09, p = .004$.

Post-hoc analyses demonstrated that Asian patients had significantly lower PHQ-9 scores at baseline ($EMM = 9.70, 95\%\ CI[9.25,10.15]$) from other patients ($EMM = 10.73, 95\%\ CI[10.45,11.00]$). This difference from Asian patients with all others persisted could be observed at weeks 2, 4, 6, 8, and 12 (see Figure 2). Interestingly, at all timepoints, there Asian patients had a lower PHQ-9 score than white patients, and this persisted even at the end of treatment. There were no other significant differences between racial/ethnic groups at each timepoint. This indicates that Asian individuals are likely to report lower PHQ-9 symptoms across some timepoints, but this difference was markedly noticeable between Asian and White patients and persistent across all timepoints.

Table 2.
ANOVA conducted on the mixed-effects model examining the interaction of Time and Race & Ethnicity on Depression scores

Effect	F	df^{S}	df^{S}_{res}	p
Time	648.53	6	17,697.09	< .001
Race & Ethnicity	7.60	3	4,866.13	< .001
Age	39.73	1	4,493.22	< .001
Gender	6.81	2	4,598.44	.001
Completion	46.74	1	5,742.73	< .001
Time × Race & Ethnicity	2.09	18	17,724.67	.004

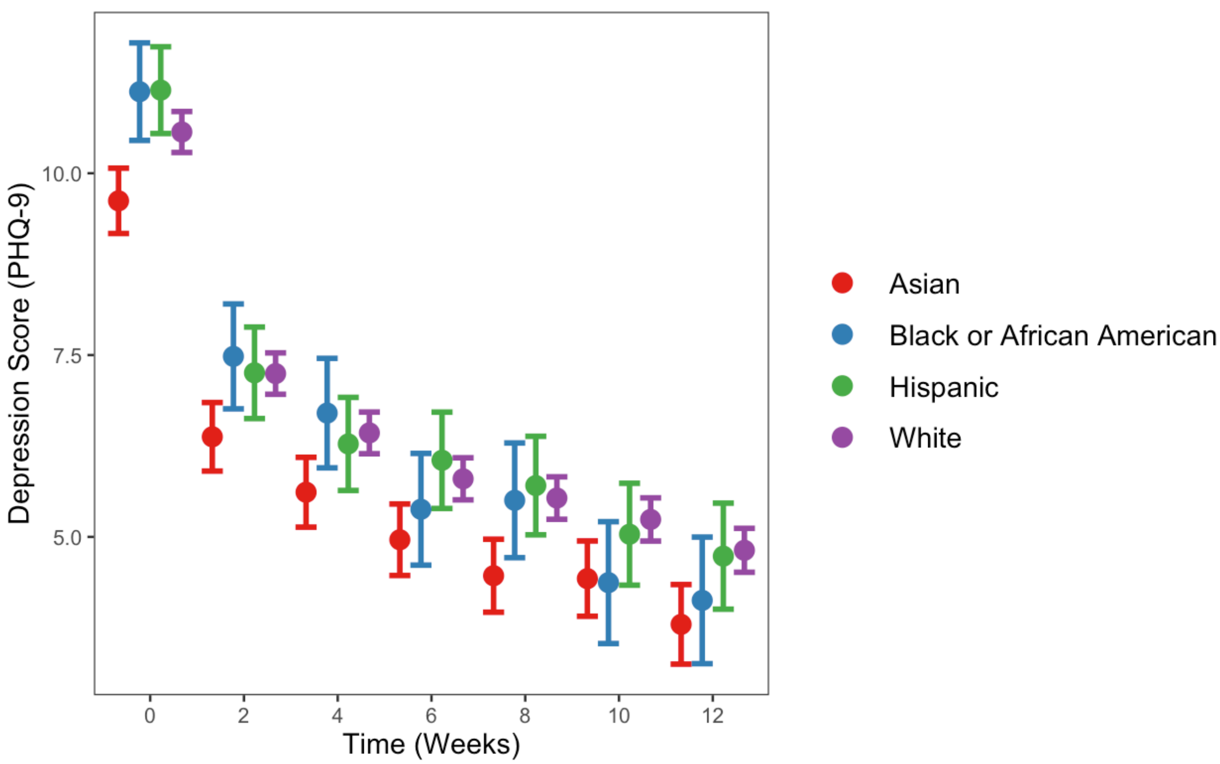

Figure 1.
Estimated marginal means of the mixed-effects model for the interaction between Time and Race & Ethnicity

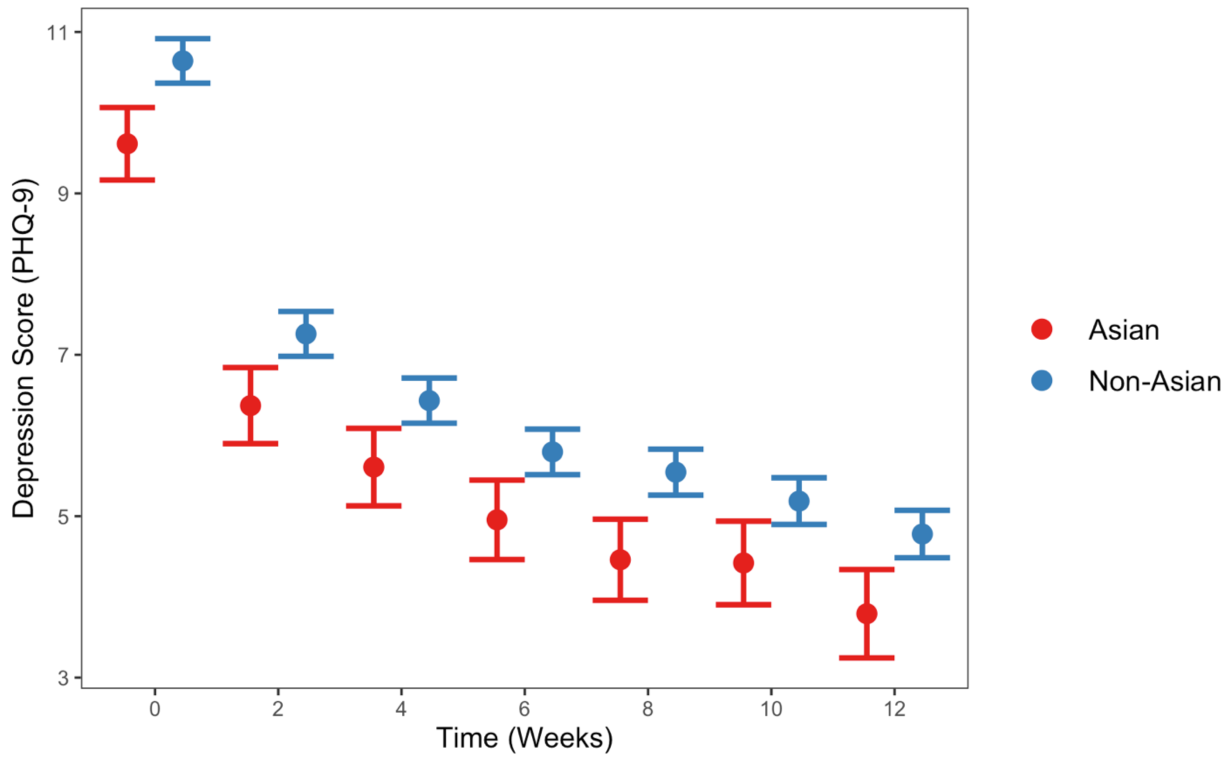

Figure 2.
Estimated marginal means of the mixed-effects model for the interaction between Time and Race & Ethnicity when contrasting between Asian and Non-Asian participants on Depression Scores.

For the GAD-7, there was a significant main effect for Race $F(3{,}4{,}905.01) = 5.78$, $p = .001$, as well as a significant main effect for Time $F(6{,}17{,}697.09) = 648.53$, $p < .001$, see Table 3. This indicates that, independent of time, there was a significant effect of race/ethnicity on GAD-7. Post-hoc analyses revealed that Hispanic individuals had significantly higher GAD-7 scores than Asian individuals, $\beta = 0.74, p = 0.009;\ 95\%\ CI\ [0.18{,}1.31]$ (see Figure 3). Additionally, there was a significant interaction between Time and Race, $F(18{,}17{,}935.21) = 2.30$, $p = .001$. Post-hoc analyses revealed that Asian individuals had significantly lower GAD-7 scores than their white counterparts at weeks 6 and 10. There were no other significant differences between racial/ethnic groups at each timepoint. Additional post-hoc analyses were run to determine if white individuals had higher GAD-7 scores at any of the timepoints recorded, compared with the other racial groups combined, but no statistically significant differences were observed.

Table 3.
ANOVA conducted on the mixed-effects model examining the interaction of Time and Race & Ethnicity on Anxiety scores

Effect	F	df^S	df^S_{res}	p
Time factor	751.25	6	17,903.08	< .001
Race & Ethnicity	5.78	3	4,905.01	.001
Gender	10.23	2	4,583.59	< .001
Age	72.07	1	4,440.23	< .001
Completion	45.34	1	6,292.50	< .001
Time × Race & Ethnicity	2.30	18	17,935.21	.001

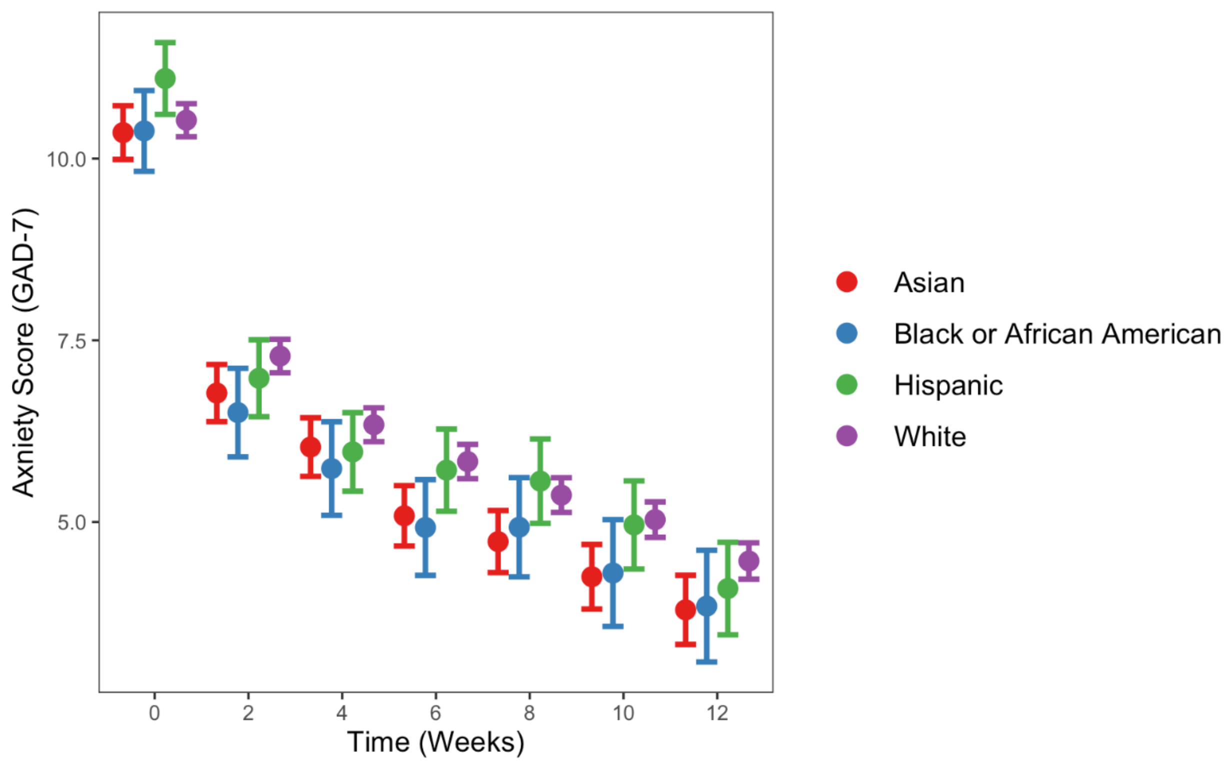

Figure 3.
Estimated marginal means of the mixed-effects model for the interaction between Time and Race & Ethnicity when contrasting between Asian and Non-Asian participants on Anxiety Scores.

Gender & Mental Health Symptoms

Similar analyses were run with gender as a predictor for PHQ-9 and GAD-7. For PHQ-9, there was a significant main effect for Time, $F(6,17{,}671.62) = 172.98$, $p < .001$ (see Table 4; see Figure 4). Additionally, there was a significant main effect for gender, $F(2,4{,}841.54) = 5.98$, $p = .003$. More specifically, non-binary individuals ($\beta = 0.26, p = 0.035, 95\% CI$ $[0.02 - 0.50]$) and women ($\beta = 0.11, p < 0.001, 95\% CI$ $[0.05 - 0.17]$) reported higher PHQ-9 scores, independent of Time. Post-hoc analyses revealed that, while there were significant main effects, there were no significant differences in estimated marginal means at any timepoint for the different genders.

Table 4.
ANOVA conducted on the mixed-effects model examining the interaction of Time and Gender on Depression scores

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Time	8,130.13	1,355.02	6	17,671.62	172.98	0.00
Gender	93.75	46.87	2	4,841.54	5.98	0.00
Age	311.24	311.24	1	4,485.95	39.73	0.00
Race & Ethnicity	195.78	65.26	3	4,634.77	8.33	0.00
Completion	371.18	371.18	1	5,733.09	47.38	0.00
Time x Gender	64.88	5.41	12	17,667.72	0.69	0.76

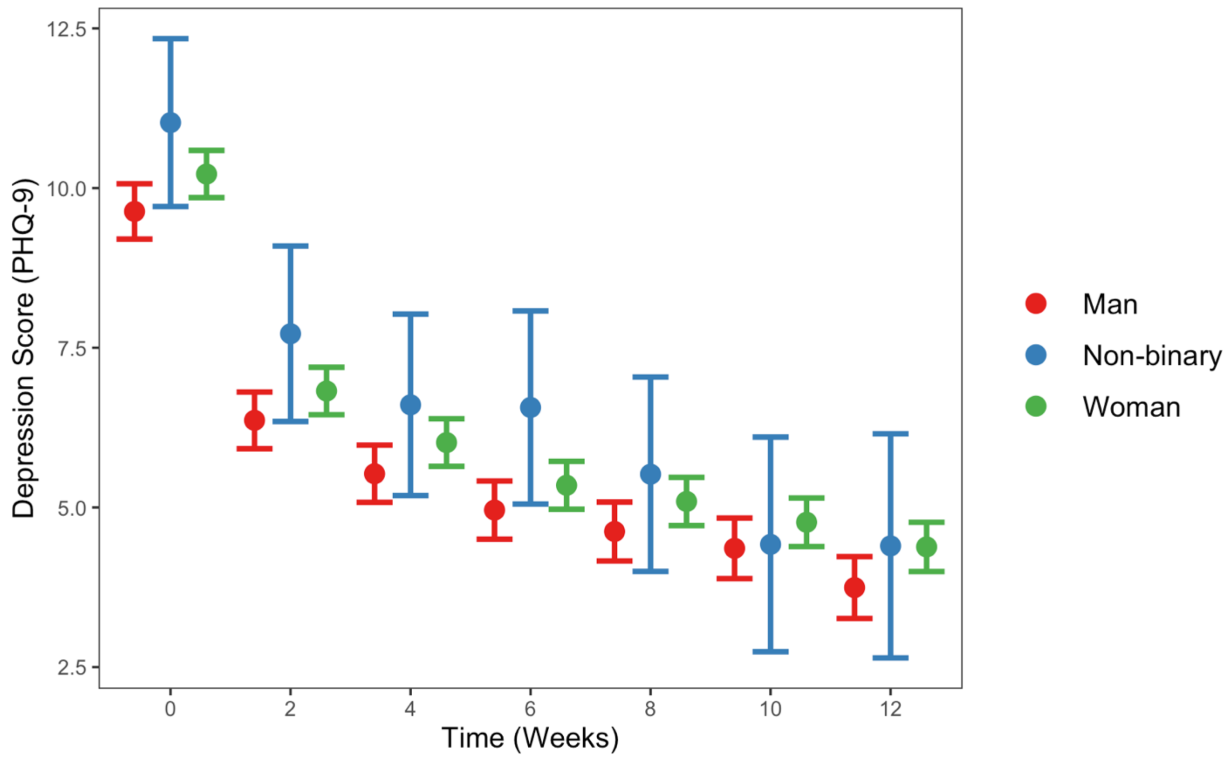

Figure 4.
Estimated marginal means of the mixed-effects model for the interaction between Time and Gender on Depression Scores.

For GAD-7, there was a significant main effect for Time, $F(6,17{,}877.31) = 166.32, p < .001$ (see Table 5; see Figure 5). Additionally, there was a significant main effect for gender, $F(2,4{,}881.32) = 9.39, p < .001$. Women ($\beta = 0.12, p < 0.001, 95\% CI$ $[0.07 - 0.18]$) reported higher GAD-7 scores, independent of Time, but not non-binary participants($\beta = -0.08, p = 0.476, 95\% CI$ $[-0.31 - 0.15]$).

Table 5.
ANOVA conducted on the mixed-effects model examining the interaction of Time and Gender on Anxiety scores

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Time	7,302.80	1,217.13	6	17,877.31	166.32	0.00
Gender	137.47	68.74	2	4,881.32	9.39	0.00
Age	526.41	526.41	1	4,432.41	71.93	0.00
Race & Ethnicity	96.07	32.02	3	4,631.17	4.38	0.00
Completion	339.16	339.16	1	6,279.50	46.35	0.00

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Time x Gender	83.12	6.93	12	17,868.36	0.95	0.50

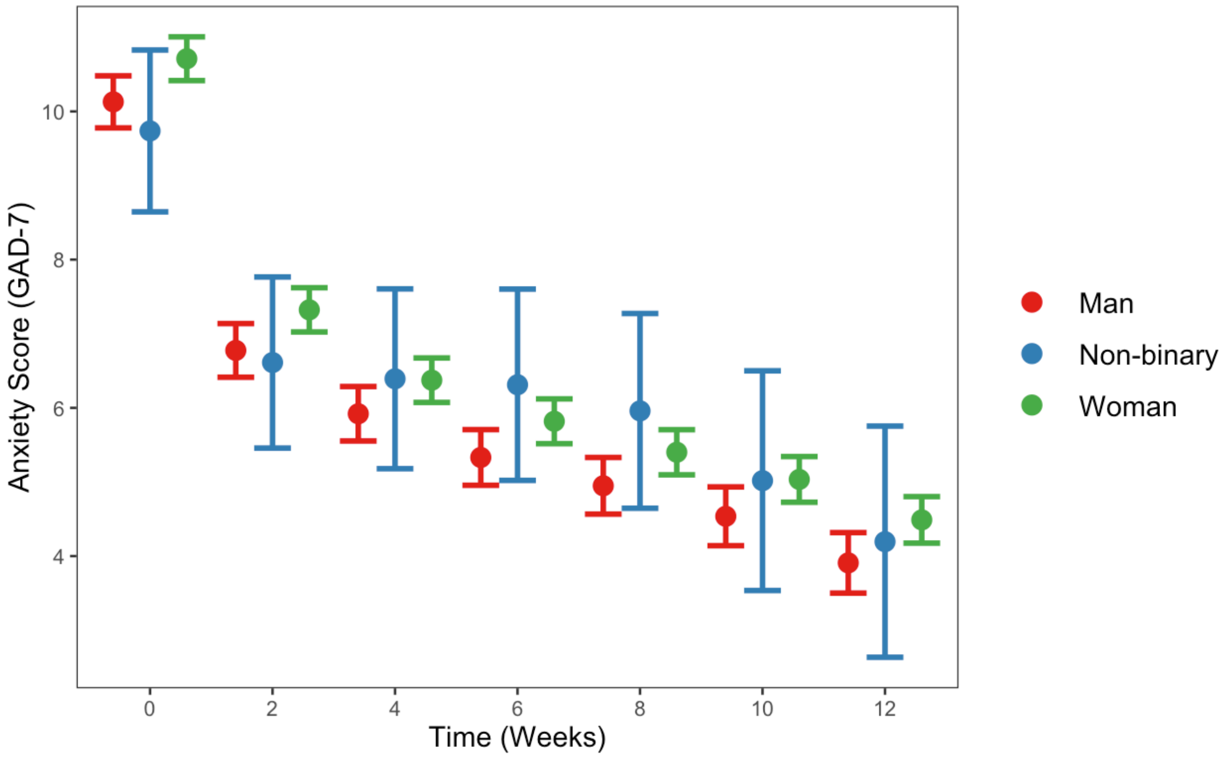

Figure 5.
Estimated marginal means of the mixed-effects model for the interaction between Time and Gender on Anxiety Scores.

Age & Mental Health Symptoms

Examining the relationship between mental health symptoms and age brackets, for PHQ-9, there was a significant main effect for Time, $F(6,17{,}478.74) = 627.62$, $p < .001$ (see Table 6; see Figure 6). Additionally, there was a significant main effect for age, $F(5{,}4{,}631.94) = 8.54$, $p < .001$. Table 7 shows the estimated marginal means for the age brackets, independent of Time. The 18-25 age group is significantly higher than any other age group. Additionally, the 26-34 group is significantly higher than all older age groups. No other groups were significantly different from one another. Figure 6 demonstrates that there are significant differences between groups within each timepoint. Most noticeably, age 18-25 was, once again, significantly higher than most age groups. What is most noticeable, however, is the clear pattern that within each timepoint, there is a downward slope for the relationship between age and PHQ-9 scores.

Table 6.
ANOVA conducted on the mixed-effects model examining the interaction of Time and Age on Depression scores

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Time	29,480.82	4,913.47	6	17,478.74	627.62	0.00
Age	334.21	66.84	5	4,631.94	8.54	0.00
Gender	101.75	50.87	2	4,587.94	6.50	0.00
Race & Ethnicity	179.18	59.73	3	4,630.57	7.63	0.00
Completion	385.89	385.89	1	5,742.75	49.29	0.00
Time x Age	287.11	9.57	30	17,580.69	1.22	0.19

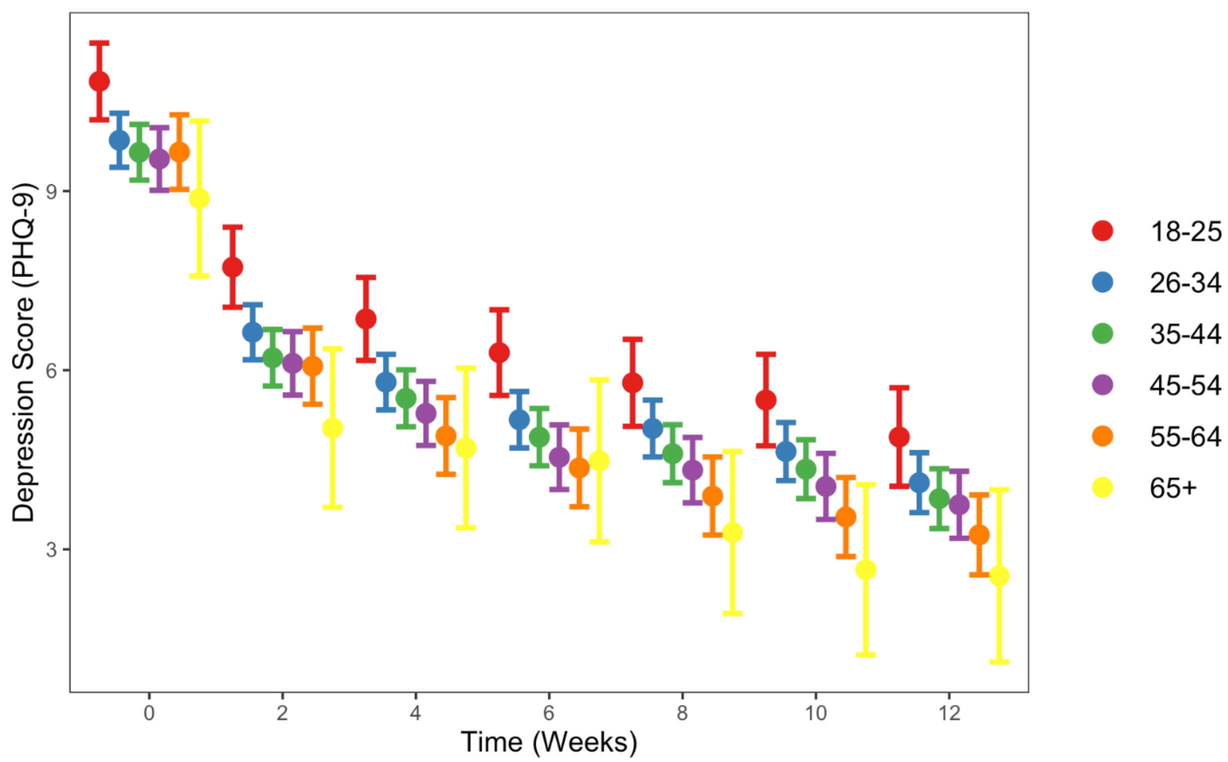

Figure 6.
Estimated marginal means of the mixed-effects model for the interaction between Time and Age on Depression Scores.

Table 7.
Estimated marginal means for age brackets, independent of Time, on Depression Scores.

Age	Estimate	95% CI
18-25	8.75	8.29 - 9.21
26-34	7.60	7.37 -7.83
35-44	7.14	6.91 - 7.37
45-54	6.92	6.63 - 7.21
55-64	6.56	6.16 - 6.96
65+	6.10	5.04 - 7.17

Examining the relationship between mental health symptoms and age brackets, for GAD-7, there was a significant main effect for Time, $F(6,17,618.90) = 741.35, p < .001$ (see Figure 7). Additionally, there was a significant main effect for age,$F(5,4,607.75) = 13.98, p < .001$. There was no significant interaction between Time and Age,$F(30,17,752.88) = 1.38, p = .079$. Table 8 shows the estimated marginal means for the age brackets, independent of Time. Similar to PHQ-9, the 18-25 age group was significantly higher than any other age group. Additionally, the 26-34 group was significantly higher than all older age groups. The 35-44 year old group and the 45-54 was significantly higher than the 55-64 and 65+ group. Figure 7 demonstrates that there were significant differences between groups within each timepoint. Most noticeably, age 18-25 was, once again, significantly higher than most age groups. Again, similar to the PHQ-9 findings, there is the clear pattern that, within each timepoint, there is a downward slope for the relationship between age and GAD-7 scores.

Table 8.
Estimated marginal means for age brackets, independent of Time, on Depression Scores.

Age	Estimate	95% CI
18-25	8.28	7.92 – 8.65
26-34	7.65	7.47 – 7.83
35-44	7.19	7.01 – 7.37
45-54	6.85	6.62 – 7.08
55-64	6.27	5.96 – 6.59
65+	5.94	5.11 – 6.89

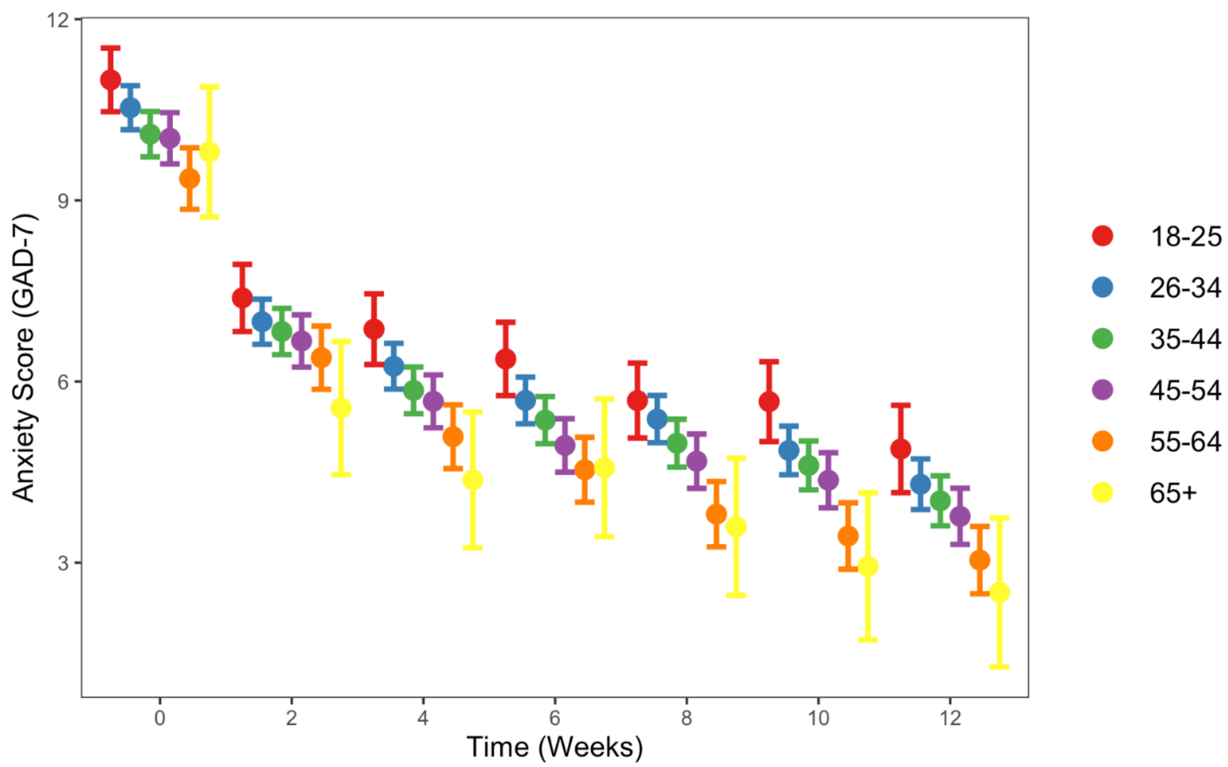

Figure 7.
Estimated marginal means of the mixed-effects model for the interaction between Time and Age on Depression Scores.

Completion Rates

In order to determine if there were differences between different racial and gender groups with regards to dropout, a logistic regression was conducted. When examining the effects of Race on completion, Black participants had a significantly lower likelihood of completing when compared with Asian participants (OR = 0.64, 95% CI [0.47, 0.89]; see Figure 8; see Table 9). Additionally Black participants had a significantly lower likelihood of completing compared to White individuals, (OR = 0.69, 95% CI [0.52, 0.91]). In contrast, there were no significant differences in completion outcomes across genders (see Figure 9; see Table 10).

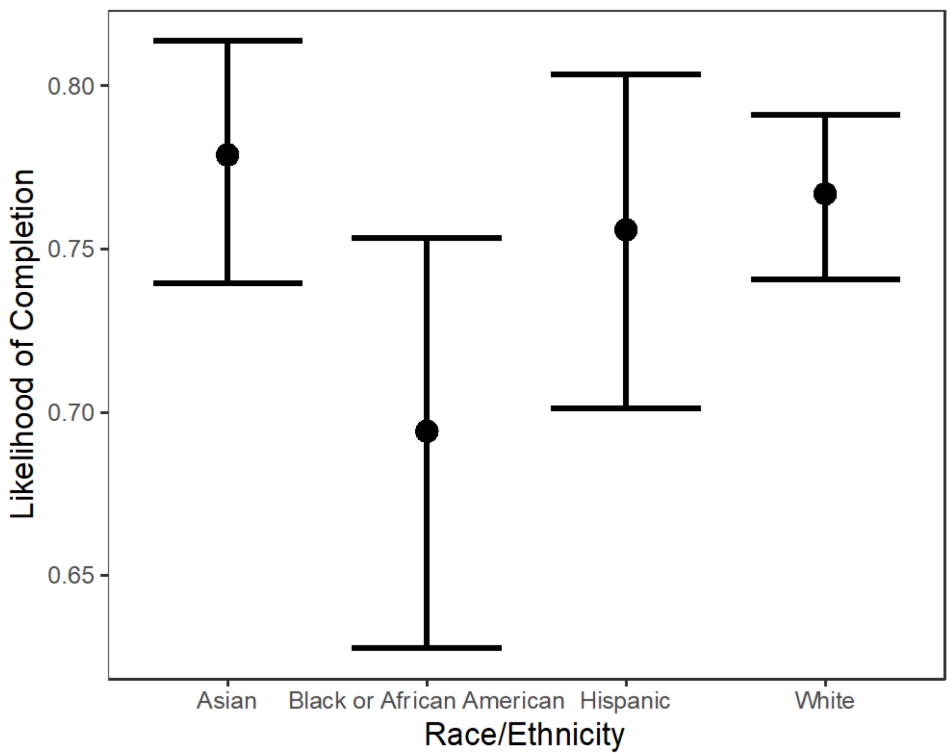

Figure 8.
Odds ratio of completing the MHP based on Race & Ethnicity.

Table 9.
Contrast of odds ratios for completion between different Races & Ethnicities.

Contrast	ΔM	95% CI	$t()$	p
Asian - Black or African American	0.44	[0.01, 0.86]	2.66	.039
Asian - Hispanic	0.13	[-0.28, 0.53]	0.81	.848
Asian - White	0.07	[-0.19, 0.33]	0.67	.909
Black or African American - Hispanic	-0.31	[-0.80, 0.17]	-1.65	.352
Black or African American - White	-0.37	[-0.74, 0.00]	-2.58	.048
Hispanic - White	-0.06	[-0.41, 0.29]	-0.45	.970

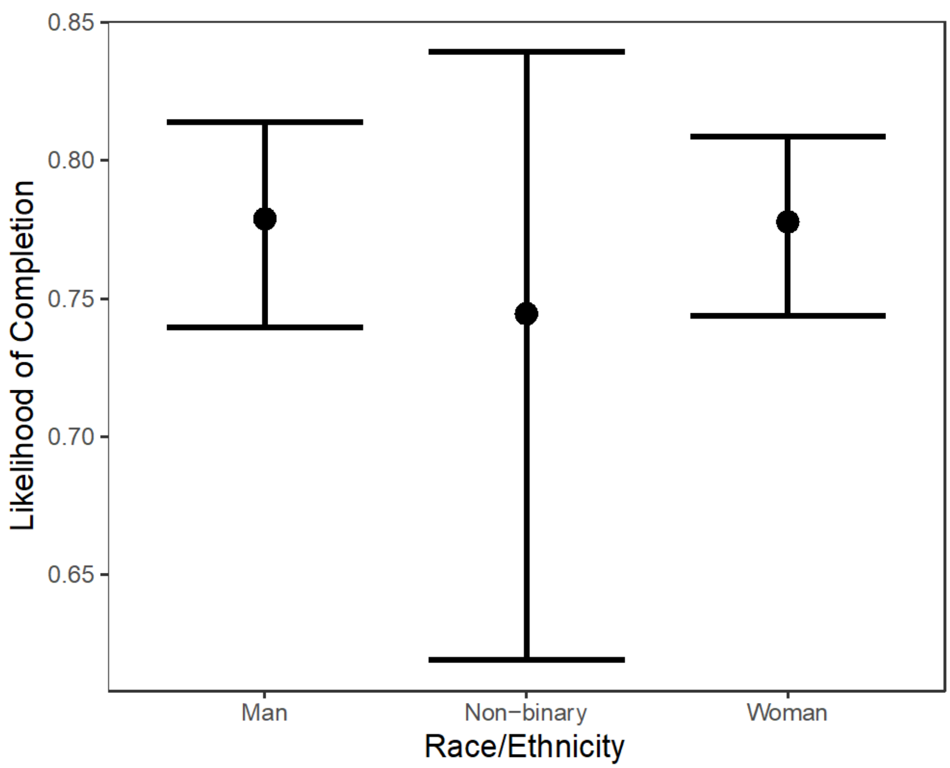

Figure 9.
Odds ratio of completing the MHP based on Gender.

Table 10.
Contrast of odds ratios for completion between different Races & Ethnicities.

Contrast	ΔM	95% CI	$t()$	p
Man - Non-binary	0.19	[-0.49, 0.87]	0.66	.789
Man - Woman	0.01	[-0.18, 0.19]	0.08	.996
Non-binary - Woman	-0.18	[-0.85, 0.48]	-0.65	.795

Study 2

Method

Data was collected from a sample of 7163 individuals who started in the MHP between Apri 27th, 2020 to February 27th, 2023. Due to this, there are 24 separate permutations for how sex and gender could combine. A participant's reported gender took precedent over their reported sex, though if their gender was not reported than their sex was used as their gender. If a participant did not answer or chose "Prefer not to disclose," then they were marked as "Other." Due to a low group size, 18 were removed from the dataset and subsequent analyses if they were marked as "Other". Any individual who indicated that they were trans, "other" or a gender other than their sex were considered "Gender Expansive." Individuals who only reported their gender or sex (but not either) were considered Cisgender, as well as anyone who reported a gender that was congruent with their sex.

Statistical Analyses

For the second part of this study, I will specifically examine the differences in suicidality between cis-gender and gender expansive individuals. In particular, I am interested to see if there are differences in the trajectories of suicidality across time in the MHP. In order to do this, I will use time as a numerical value and have gender as a moderating variable for this to determine if the slopes in reduction of suicidal symptoms are significantly steeper than in cis-gender clients, accounting for baseline PHQ-8 scores (PHQ-9 without the suicide-related question).

Results

See Table 11 for descriptive statistics. In order to determine if there were significant differences in suicidality at baseline between cisgender and gender expansive participants, I ran a linear model with the 9th item of the PHQ-9 as the outcome and gender as a predictor, with age and baseline PHQ-8 symptoms as covariates. The effect of Gender is statistically significant and positive (b = 0.27, 95% CI [0.16, 0.37], $t(5146) = 4.95$, $p < .001$; see Figure 10). This indicates that, even controlling for demographic and clinical characteristics, gender expansive participants have significantly higher levels of suicidality than cisgender participants.

Table 11.
Summary of descriptive statistics for the participants used in analyses.

Characteristic	Overall, N = 5,154[1]	Cisgender, N = 5,052[1]	Gender Expansive, N = 102[1]	p-value[2]
Completion				0.2
completer	3,844 / 5,154 (75%)	3,774 / 5,052 (75%)	70 / 102 (69%)	
drop-out	1,310 / 5,154 (25%)	1,278 / 5,052 (25%)	32 / 102 (31%)	
Age	39 (11)	39 (11)	35 (10)	<0.001
Race/ethnicity				0.022
Asian	636 / 5,154 (12%)	629 / 5,052 (12%)	7 / 102 (6.9%)	
Black or African American	246 / 5,154 (4.8%)	243 / 5,052 (4.8%)	3 / 102 (2.9%)	
Hispanic	695 / 5,154 (13%)	683 / 5,052 (14%)	12 / 102 (12%)	
Other	67 / 5,154 (1.3%)	62 / 5,052 (1.2%)	5 / 102 (4.9%)	
White	3,510 / 5,154 (68%)	3,435 / 5,052 (68%)	75 / 102 (74%)	
Baseline PHQ-9	11 (6)	11 (6)	12 (6)	0.2
Baseline GAD-7	11 (5)	11 (5)	11 (5)	0.8
PHQ item 9				<0.001
0	4,188 / 5,154 (81%)	4,118 / 5,052 (82%)	70 / 102 (69%)	
1	747 / 5,154 (14%)	729 / 5,052 (14%)	18 / 102 (18%)	
2	138 / 5,154 (2.7%)	133 / 5,052 (2.6%)	5 / 102 (4.9%)	
3	81 / 5,154 (1.6%)	72 / 5,052 (1.4%)	9 / 102 (8.8%)	

[1]n / N (%); Mean (SD)

[2]Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test

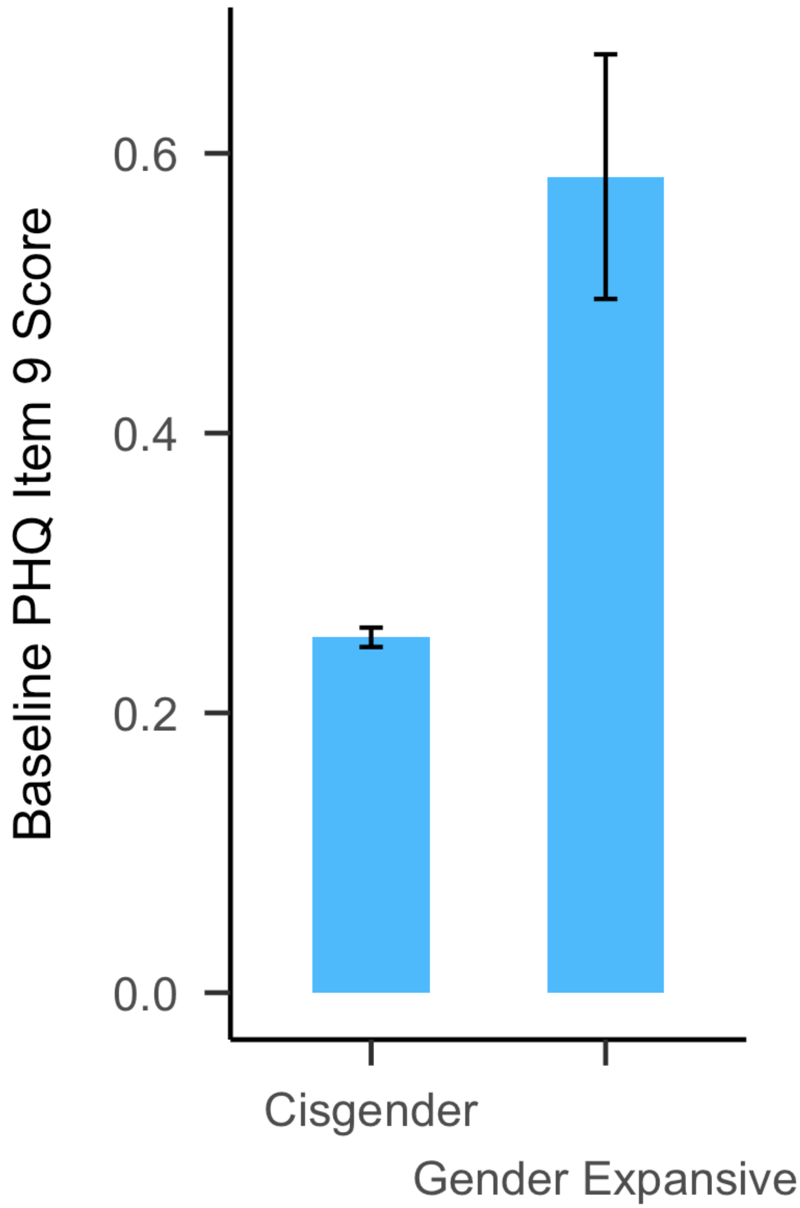

Figure 10.
Estimated marginal means from a linear model with the 9th item of the PHQ-9 as the outcome and gender as a predictor, with age and baseline PHQ-8 symptoms as covariates.

Before I examined how the 9th item of the PHQ-9 changed over time, I first examined the trajectories of non-suicidal symptoms from the PHQ-9 (in other words the first 8 items of the PHQ-9, or the PHQ-8). Table 12 demonstrate that there was a significant main effect for Time, where PHQ-8 symptoms reduced over time. There was also a significant main effect for gender, indicating that gender expansive participants reported higher overall scores on the PHQ-8, independent of time. Interestingly, there was no significant interaction effect (as can be seen in Figure 11), indicating that the rate at which these symptoms are decreasing is roughly the same regardless of gender.

Table 12.
Results of a mixed-effects model that examined the interaction of Gender and Time on PHQ-8 scores.

	PHQ-8					
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*std. p*
(Intercept)	10.44	-0.02	10.04 – 10.83	-0.04 – 0.00	**<0.001**	0.063
Gender [Gender Expansive]	0.79	0.18	-0.04 – 1.63	0.03 – 0.33	0.063	**0.017**
Time	-0.44	-0.32	-0.45 – -0.43	-0.33 – -0.32	**<0.001**	**<0.001**
Age	-0.03	-0.06	-0.04 – -0.02	-0.08 – -0.04	**<0.001**	**<0.001**
Completion [Drop-out]	1.87	0.35	1.63 – 2.12	0.31 – 0.40	**<0.001**	**<0.001**
Gender [Gender Expansive] × Time	0.04	0.03	-0.04 – 0.11	-0.03 – 0.08	0.351	0.351
Random Effects						
σ^2	8.74					
$\tau_{00\ userID}$	16.39					
ICC	0.65					
N $_{userID}$	7163					
Observations	33341					
Marginal R^2 / Conditional R^2	0.140 / 0.701					

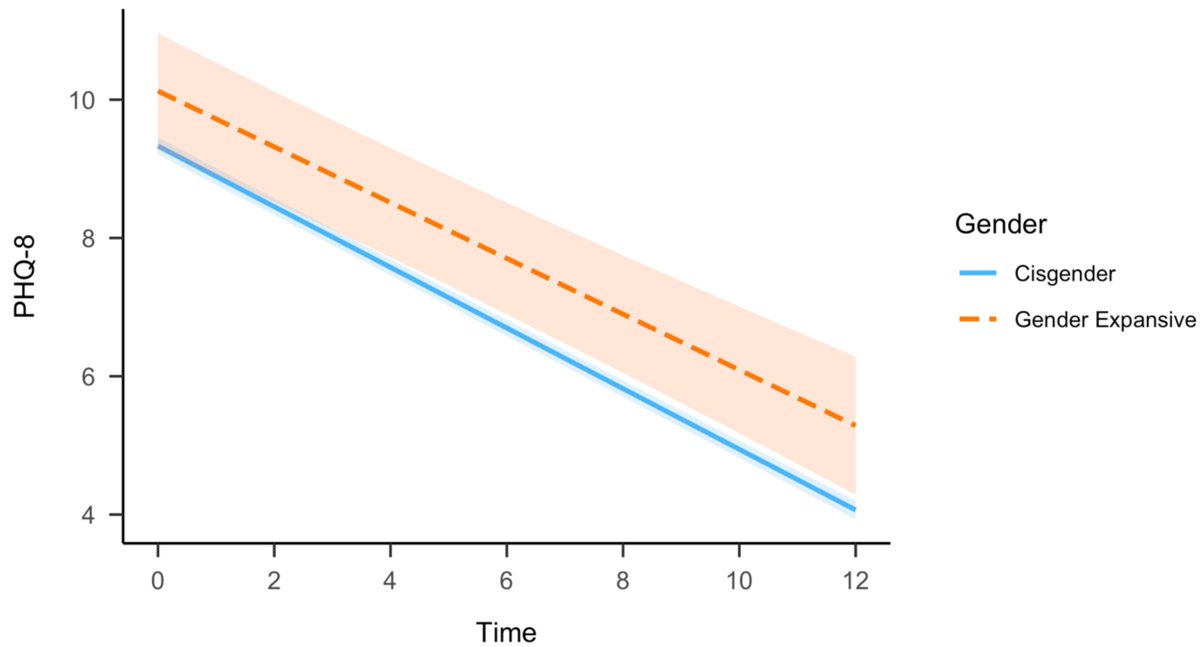

Figure 11.
Interaction plot depicting the (non-significant) interaction between Time and Gender on PHQ-8 scores.

Next, I wanted to determine if the rate of change for the 9th item of the PHQ-9 across the MHP was significantly different. In order to do this, I ran another mixed-effect model; Table 13 provides the results for this model. Results indicate that there was a significant main effect for time, such that suicidality went down throughout the duration of the MHP. Additionally, there was a significant effect for gender, such that gender expansive participants reported higher suicidality than cisgender participants. There was, however, a significant interaction effect, indicating that the rate at which suicidality decreased for gender expansive participants was greater than the rate that cisgender participants decreased (see Figure 12).

Table 13.
Results of a mixed-effects model that examined the interaction of Gender and Time on PHQ item 9 scores.

	PHQ item 9					
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*std. p*
(Intercept)	0.26	-0.02	0.23 – 0.29	-0.04 – 0.00	**<0.001**	0.062
Gender [Gender Expansive]	0.29	0.49	0.22 – 0.36	0.34 – 0.65	**<0.001**	**<0.001**
Time	-0.01	-0.12	-0.01 – -0.01	-0.12 – -0.11	**<0.001**	**<0.001**
Age	-0.00	-0.06	-0.00 – -0.00	-0.08 – -0.04	**<0.001**	**<0.001**
Completion [Drop-out]	0.07	0.18	0.05 – 0.09	0.13 – 0.22	**<0.001**	**<0.001**
Gender [Gender Expansive] × Time	-0.02	-0.18	-0.03 – -0.01	-0.25 – -0.11	**<0.001**	**<0.001**
Random Effects						
σ^2	0.08					
$\tau_{00\ userID}$	0.10					
ICC	0.53					
N_{userID}	7163					
Observations	33342					
Marginal R^2 / Conditional R^2	0.028 / 0.548					

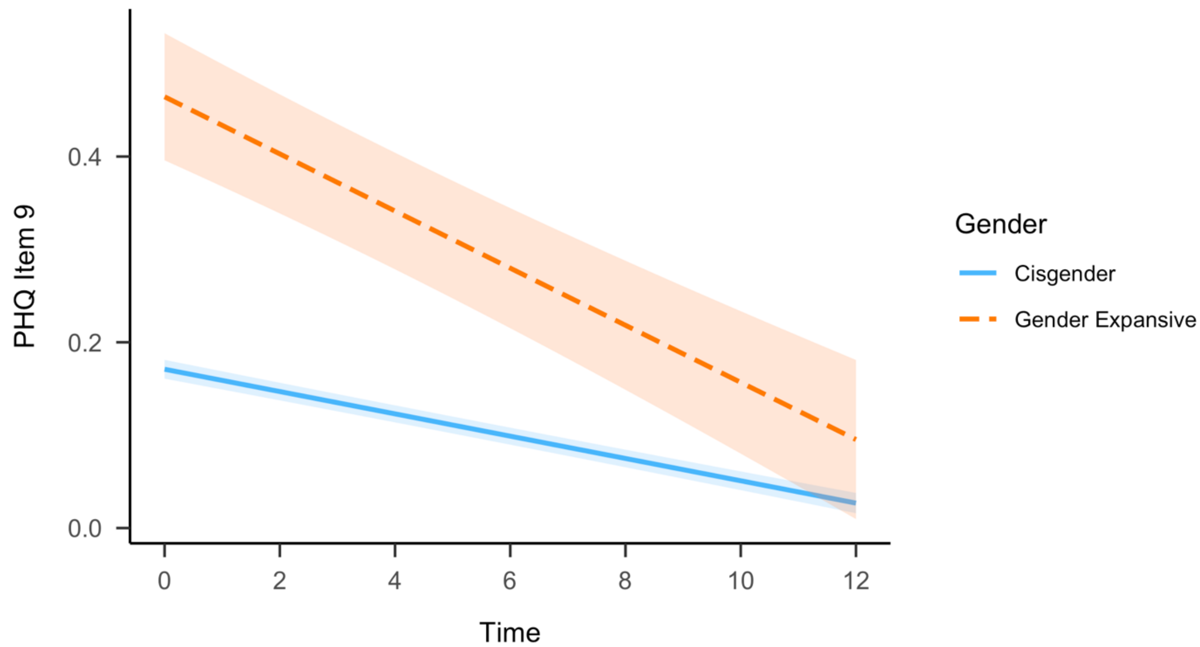

Figure 12.
Interaction plot depicting the interaction between Time and Gender on PHQ item 9 scores.

In order to determine if these results occurred over and above the reductions in non-suicidal PHQ-9 symptoms (i.e., PHQ-8), I added the PHQ-8 scores as a covariate to the above model. Interestingly, all aspects of the model stayed the same (see Table 14); however, as can be seen in Figure 13, the slope for cisgender individuals became positive. A simple slope analyses revealed that cisgender individuals had a significantly, but slight positive slope of 0.0010 (*SE* = 0.0005). This indicates that, when accounting for PHQ-8 symptoms, all other reductions in suicidality are removed for cisgender participants. This is not true for gender expansive participants, however, as they still maintained a negative slope of –0.02 (*SE* = 0.0035). This indicates that, over and above the reductions that can be attributed to reductions in PHQ-8 symptoms, gender expansive people still experience significant reductions in suicidality throughout the MHP.

Table 14.
Results of a mixed-effects model that examined the interaction of Gender and Time on PHQ item 9 scores, accounting for PHQ-8 scores

	PHQ item 9					
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*std. p*
(Intercept)	-0.05	-0.01	-0.08 – -0.02	-0.03 – 0.01	**0.002**	0.222
Gender [Gender Expansive]	0.27	0.42	0.21 – 0.33	0.28 – 0.56	**<0.001**	**<0.001**
Time	0.00	0.01	0.00 – 0.00	0.00 – 0.02	**0.033**	**0.033**
Age	-0.00	-0.04	-0.00 – -0.00	-0.06 – -0.02	**<0.001**	**<0.001**
completion [Drop-out]	0.02	0.04	-0.00 – 0.04	-0.00 – 0.09	0.072	0.072
PHQ 8	0.03	0.38	0.03 – 0.03	0.37 – 0.40	**<0.001**	**<0.001**
Gender [Gender Expansive] × Time	-0.02	-0.19	-0.03 – -0.01	-0.26 – -0.12	**<0.001**	**<0.001**
Random Effects						
σ^2	0.08					
$\tau_{00\ userID}$	0.08					
ICC	0.50					
N_{userID}	7163					
Observations	33341					
Marginal R^2 / Conditional R^2	0.146 / 0.576					

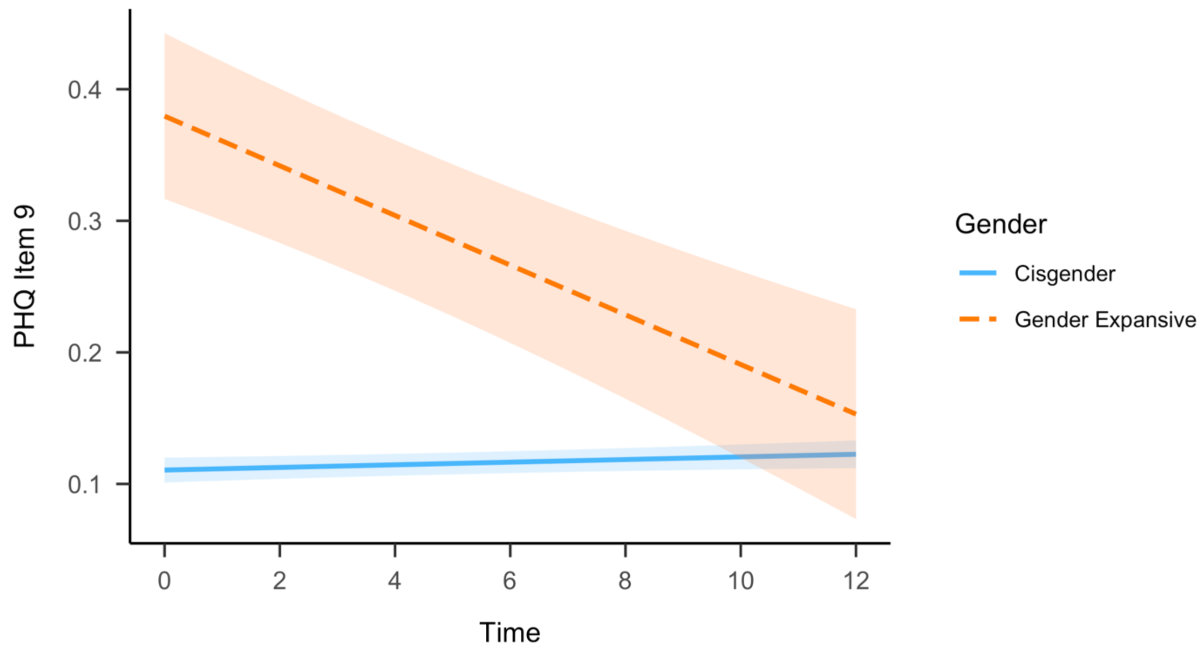

Figure 13.
Interaction plot depicting the interaction between Time and Gender on PHQ item 9 scores, accounting for PHQ-8 scores.

Additional sensitivity analyses were conducted by adding in Race & Ethnicity as a covariate. This was not included in the above analyses, because these questions were added later. Many individuals did not have data for this, so it reduced the amount of individuals within the analyses. Despite this, there were no differences in the observed effects for any of the models, adding to the robustness of these findings.

Discussion

In study 1, I found that there was a significant main effect for time, such that as time increased, depression and anxiety scores decreased, independent of race and ethnicity. For each racial and ethnic group, there were significant decreases in depression and anxiety scores, indicating that this program was associated with reductions in mental health symptoms, regardless of race or ethnicity. When examining differences across time-points, Asian participants had a significantly lower starting depression score than all other races and ethnicities combined into one group, and that persisted throughout the duration of the intervention. Additionally, when comparing Asian participants to specific groups, Asian individuals only had a significant difference from White participants, which persisted across all timepoints. For anxiety scores, results showed that Hispanic participants had higher anxiety scores than Asian participants, independent of time. Additionally, at multiple timepoints throughout the program, Asian participants had lower anxiety scores than White participants at 2 timepoints, but those differences were not present at the beginning or end of treatment. In broad strokes, it appears that Asian individuals had lower scores than White individuals, especially regarding depression scores, but for all other racial and ethnic groups, there were no consistent differences in scores across time.

Regarding depression and anxiety scores across genders, our results demonstrated that women have higher depression and anxiety scores, independent of timepoint—however, there was on singular timepoint in which there was a significant difference between any gender group. Regarding age, we found that across both depression and anxiety scores, older individuals were likely to report lower mental health symptoms than younger individuals. Regarding completion rates, Black participants were less likely to complete the intervention than Asian or White participants. There were no significant differences in completion rates across the different gender groups.

For study 2, we wanted to examine differences in treatment trajectories based on whether a participant was cisgender or gender expansive. Initially, we examined PHQ-8 scores across time (independent of the 9th item of the PHQ-9—the suicidality question), and we found that, though gender expansive individuals had higher overall PHQ-8 scores, there was almost an identical negative slope across time, indicating significant reductions in depressive symptoms independent of suicidality. Regarding the 9th item of the PHQ-9, we found that gender expansive participants had significantly higher baseline suicidality than cisgender participants. We then ran a mixed-effects model to see how suicidality changed across the intervention (not accounting for PHQ-8 scores), and the results showed that there was a significant interaction, such that, despite having higher suicidality throughout the program, they still had significantly more reductions in suicidality than their cisgender counterparts. However, we wanted to ensure that these reductions in suicidality were not accounted for by other reductions in PHQ-8 scores (which would indicate that suicidality would be reduced proportionally to other symptoms). What we found was that, when accounting for other symptoms, the reductions in suicidality vanished in cisgender participants, indicating their reductions in suicidality were proportional to their reductions in other symptoms. Most interestingly, though, the reductions in suicidality for gender expansive participants remained significant, indicating that these reductions can be explained over and above simply their reductions in other depressive symptoms.

Unfortunately, research examining the relationship between Asian individuals and depression and anxiety scores within a therapy context is limited, especially within DMH interventions; however, there have been some studies. One such study examined the relationship between psychological well-being between Chinese and American individuals and found that there were no differences in depression scores between the two groups (J. Zhang & Norvilitis, 2002). Importantly, this study did not focus on a population of those receiving therapy treatment, making the sample between this study and theirs different enough that comparisons should be made with caution. In another study among different Asian populations in Canada, researchers found that other Asian participants were less likely than White individuals to have had any mood and anxiety disorder in their lifetime (Tiwari & Wang, 2006). Though, again, this study did not examine individuals that were seeking treatment for their mental health issues. Another study had similar findings in an Australian population, such that Asian individuals were likely to score lower on depression and anxiety scores than non-Asian individuals (Comino et al., 2001). While these findings are mixed, there does seem to be some evidence that Asian individuals under-report their symptoms and could be one mechanism for why our study found lower depression scores for Asian participants in our sample. Additionally, Asian individuals are likely to report more somatic symptoms over emotional symptoms (Edman et al., 1999; Novick et al., 2013). If Asian individuals were under-reporting their non-somatic symptoms on the PHQ-9, this could also lead to the lower scores observed in our study.

With regards to the findings that older age is related to lower depression and anxiety scores, this is consistent with other studies. One recent study found that age was a positive predictor in outcomes for Cognitive Behavioral Therapy (CBT), indicating that older age is related to reduced mental health symptoms (von Brachel et al., 2019). Additionally, our findings that women are more likely to report higher depression and anxiety scores are also consistent with other research. Research has consistently shown that there is a higher prevalence of depression in women compared to men (Hyde et al., 2008; Parker & Brotchie, 2010).

For study 2 our findings regarding higher starting suicidality in gender expansive clients are consistent with other research studies. One recent study found that gender minorities were more likely to report higher suicidality than non-gender minorities (Progovac et al., 2020). To our knowledge, no one has examined the trajectory of gender expansive suicidality symptoms within a digital mental health context. Heck and colleagues (2017) conducted a systematic review that determined that there was a severe lack of reporting for gender and sexual minorities in RCTs for mental health treatment, despite this being such a critically vulnerable population. More specifically, of the 232 articles that they reviewed, only 1 reported participant's sexual orientation and no study reported participants' gender identity. This is despite the high prevalence of mental health issues in these populations. One study found that in a sample of 130 transgender participants, a third had sought treatment for mental health issues, but a significant portion of those were unable to receive services. The researchers concluded that they experienced barriers to accessing mental health services, including the cost of treatment, previous bad experiences with healthcare, fear of treatment, and stigma concerns. This work has significant implications for mental health treatment within the LGBTQ+ population, as these individuals are often scared to seek out treatment. DMHI offers an alternative that can reduce some of the barriers to access that this population usually faces when seeking treatment. Additionally, our findings indicate that this particular DMHI is related to reductions in suicidality in gender expansive clients, over and above the reductions they experienced in non-suicidal depression-related symptoms.

These studies are not without their limitations. The first major limitation is the lack of a control group that is not going through the intervention. This limitation keeps us from making causal inferences with our data, so we are not able to determine whether the reductions in depression and anxiety that were observed are related to the intervention or some other factor not measured. Future studies should include control groups that are not undergoing the intervention would increase our ability to make inferences about the efficacy of the DMHI. Additionally, while the sample was relatively large for a psychological study, some of the sub-groups of participants had a relatively small size (e.g., gender expansive participants). Though mixed-effects modeling is robust enough to handle these group size imbalances, it still does not mitigate some of the biases that can arise from such a disparity in group size. In the future, outreach to these specific population for recruitment can help mitigate these biases. Additionally, many of the individuals going through the program were referred through either their employer or their insurance company. Individuals with jobs that will pay for this program or health insurance are potentially more likely to have a higher Socioeconomic Status (SES) than others, limiting the ability for us to draw inferences to a lower SES population. Future research should incorporate reduced cost options and recruitment in marginalized and lower SES populations.

These studies represent a novel addition to the literature. In particular, there is scant research literature regarding outcomes in these different populations within the context of DMHIs. We found that Age, Race, and Gender are significantly related to outcomes within the intervention. Additionally, though gender expansive individuals are more likely to report higher suicidality, they seem to see improvements in suicidality related to the mental health intervention, which is something that has not been previously explored before. Our results demonstrate that both traditional and digital mental health interventions should consider how these factors contribute to their outcomes and how they can target these populations with their interventions.

Chapter 3

Introduction

Once individuals have begun their participation in a DMHI, one of the difficulties is keeping them engaged. For example, one recent review of the literature found that treatment adherence could range wildly (from 6% to 100%) (Andrews et al., 2018). Another review points out that there are many barriers that make DMHIs more difficult to engage with than traditional therapies (Borghouts et al., 2021). These can include personality factors, such as extraversion being related to lower engagement, but neuroticism being related to higher engagement (Borghouts et al., 2021). Additionally, there can be program-related barriers, such as a lack of guidance from either automated or human therapist generated reminders. Finally, they also point out technological related barriers, such as difficulties navigating the app or glitches within the platform. Because of this, DMHIs can have a lower adherence than traditional in-person therapy (Kaltenthaler et al., 2008). In order to prevent this from happening, one of the suggested ways is to have a more tailored experience for the individual to feel connected with the platform. In order to do that, it is important for these DMHIs to know who needs to be targeted for early dropout to make sure that these individuals can be targeted by either the platform itself or the human-therapists that interact with them.

Being able to determine who might disengage from a program could lead to better, more tailored interventions that will reduce dropout and increase the overall efficacy of the program to the widest range of individuals. However, often with traditional therapy services, therapists must rely on client self-report and their own intuition in order to predict engagement with their services, which may not be indicative of their ability to complete the intervention. This is one of the major drawbacks of any self-report measure, as self-report is often only weakly to moderately related to actual behaviors and outcomes (Dang et al., 2020). One of the strengths of using these digital platforms, however, is that real-time unbiased behavioral data can be collected, which has the potential to be a much better predictor of adherence to the intervention. Additionally, with the data gathered, I am able to are able to ascertain whether or not objective measures, as measured through engagement with the app, are a better predictor of completion of a treatment program than self-reported measures, such as motivation. The aim of the studies outlined below is to determine if there are observable factors that can warn us if a potential individual is at-risk for dropping out of the program (in the hopes that it could then be used for an early intervention to prevent this). I have conducted two separate studies to accomplish this, examining engagement data, clinical scores at baseline, and chat messages between therapist and clients.

For the first study, I utilized a mixed-methods approach to determine how clients were engaging in chat messages with their therapist. I conducted a qualitative analysis on chat messages between therapist and client to see which themes were most common within these interactions. No specific hypotheses were generated for this portion of the study. For the quantitative analyses, as a part of the initial project, I read through the messages in order to generate hypotheses about what I was observing in the data. Two specific hypotheses emerged while reading the chat messages: 1) therapists who sent longer messages received fewer text messages from clients than therapist who sent shorter messages, 2) messages with questions contained within longer messages were less likely to receive a response than messages with questions contained in a shorter message. For the third part of the study, I used machine learning techniques in order to determine if there were specific topics within the chat messages that were related to completion of the MHP.

For the second study, the goal of this study was to determine which engagement variables taken within the first week MHP, would be predictive of program completion rates. I hypothesized that engagement measures (i.e., Active Days, Days Active in Chat, Time Spent Meditating, Time Spent in HRVB) would be a better predictor of program completion rates than self-reported measures (i.e., depression scores, and anxiety scores). No other specific hypotheses were made for this study.

Study 1

Method

Qualitative Analyses

Qualitative analyses were conducted on messages collected through the Meru Health Program between therapist and the client. Messages were run through an algorithm that removed any Personally Identifiable Information (PII) and Personal Health Information (PHI), such as names, dates, contact information, etc. Additionally, due to the sensitive nature of this chat data between provider and patient, I was the only one given permission to read the contents of the messages for this project. For this particular analysis, chat messages from 100 participants were randomly selected from the dataset, and the chats were subset to the messages that occurred during the first week of the program. The messages were then further split up into two separate datasets, in order to conduct the analyses in two separate steps. The first set, a "training" dataset, consisted of messages from 50 participants. I read through that dataset, and recorded recurring themes that I could use for coding. Table 15 shows the following were the 14 codes that were created:

Table 15.
Table consisting of the codes used in the qualitative analysis and the definitions for each code.

Code	Definition
Simple Pleasantries	Client having simple conversation with the therapist (e.g., "Hello")
Response to Q from therapist (Neutral)	Client responding to a banal question from the therapist
Discussing positive aspects of intervention	Client discussing aspects of the intervention that indicate a positive benefit from the program
Discussing neutral aspects of the intervention	Client discussing aspects of the intervention that do not indicate any positive or negative experiences
Discussing difficulties with the intervention practices	Client discussing issues that have arisen *specifically* related to the therapeutic intervention practices in the program
Tech difficulties	Client discussing issues that have arisen *specifically* related to the technology in the program
Indicating busy-ness	Client indicating that they have very little time because they have other obligations
Concerns about the program	Clients expressing larger concerns about the program (e.g., no face-to-face sessions with therapist)
Requesting help	Client is requesting help due to a process that has gone awry
Therapy interaction	Interaction that indicates some level of therapeutic process is occurring (e.g., psychoeducation, behavioral activation)
Discussing situations and feelings	Client discussing a situation that has arisen in their life outside of the program and any associated feelings
Conceptual Questions	Client asking questions about a specific process in the program (e.g., how does HRV-B work)
Logistics Questions	Client asking questions about something logistic within the program (e.g., how to get to a certain module)
Asking advice	Client asking for direct advice about how to accomplish a task that isn't specifically therapy related

Once these codes were created, I consulted with the chair of my dissertation committee to discuss these codes and how they would be applied. We devised a strategy that any messages I encountered that were ambiguous would be appropriately re-worded (while still maintaining the original meaning) and would be discussed with advisor to properly code it. The second step included using the Qualitative Analysis package "MAXQDA2022" (VERBI Software, 2022) to code the messages from the "testing" dataset consisting of messages from another 50 participants. Once completed, there were certain categories that contained fewer than 5 messages (e.g., “Asking advice,” “Conceptual Questions,” “Concerns about the program.”). In order to conduct subsequent simple quantitative analyses on these codes related to program outcomes, my advisor and I placed the codes into one of two categories: facilitating therapy processes, and impeding therapy processes. Table 16 depicts how the codes were categorized:

Table 16.
Table describing how each code was either placed into a “facilitating therapy processes” category or “disrupting therapy process” category.

Facilitating therapy processes	Disrupting therapy processes
Simple Pleasantries	Discussing difficulties with the intervention practices
Response to Q from therapist (Neutral)	Tech difficulties
Discussing positive aspects of intervention	Indicating busy-ness
Discussing neutral aspects of the intervention	Concerns about the program
Therapy interaction	Requesting help
Discussing situations and feelings	
Conceptual Questions	
Logistics Questions	
Asking advice	

Statistical Analyses

In order to determine whether these hypotheses were supported by the data, I conducted generalized linear modeling on the chat messages. To process the data for the first hypothesis, each therapist’s message was given a character count based on the number of characters present in the message (to be used as the outcome variable). Additionally, a variable was created for each individual client that counted the amount of chat messages they sent throughout the duration of the MHP to their therapist (to be used as the predictor variable). The ‘glmer’ function from the *lme4* package was used with a Poisson link function to specify a Poisson generalized linear regression, as the was not zero-inflated). Message length was scaled for better model fit.

Additionally, the model accounted for therapist as a random variable, to account for non-independence in the data, as there were multiple data points for each therapist.

For the second hypothesis, therapist messages were reduced to messages that had a "?" to indicate there was a question in the message. Then, any subsequent messages sent from the therapist within 5 minutes of the original message was removed, in case the therapist sent multiple messages in a brief span of time. A binary variable was created that accounted for whether or not the message following the question from the therapist was from the client or the therapist at a later time. If the following message was from the client, this was considered a "response" to the therapist's question. If the following message was from the therapist at a later time, this was considered "no response" to the therapist's question. This binary variable was then used as the outcome variable in a generalized linear model with therapist message length (containing a question) as the predictor variable. Message length was scaled for better model fit. The model accounted for clients nested within therapists, as there were repeated measures from both therapists and clients, and therapists were given multiple clients but not vice versa.

Machine Learning

In order to prepare the text data for the machine learning algorithm, the text data needed to be cleaned and preprocessed. Names were scrubbed from all text data by using a list of common first and last names as a reference—any word that matched the names on that list was removed from the messages. Next, all punctuation marks were removed from the text data. Finally, because there were many messages from the therapists that were duplicates (i.e., some therapists would copy and paste a message and send it to all their clients), all of the messages were put through a cosine similarity function that creates a matrix that calculates the similarity between all of the messages. If there were messages that were too similar (threshold > 0.9), those messages were removed from the analyses. Each message was then converted into a “document,” and fed into the structural topic model that examined the prevalence of topics within and across all of these documents. For this analysis, the number of topics was not predefined and the number of topics is set by a spectral algorithm (Mimno & Lee, 2014). The max number of iterations was set to 10,000. It was then specified within the model to estimate which topics were significantly related to completion (while adding in therapist ID as a covariate). Two separate models were ran, one containing messages from therapists and one containing messages from participants.

Results

Qualitative Analyses

Participants

One hundred participants were randomly sampled from the larger training set, with 50 in each the training and the testing dataset. In the training dataset, there were 116 messages sent from the client, and in the testing dataset there were 133 messages sent from the client. Table 17 shows the descriptive statistics of the sample used in these analyses. There were no significant differences observed between the testing and the training dataset.

Table 17.
Summary of descriptive statistics for the participants used in the qualitative analyses.

Characteristic	**Overall**, N = 100[1]	**Testing**, N = 50[1]	**Training**, N = 50[1]	**p-value**[2]
Completion				0.2
Completer	81 / 100 (81%)	38 / 50 (76%)	43 / 50 (86%)	
Drop-out	19 / 100 (19%)	12 / 50 (24%)	7 / 50 (14%)	
Gender				0.5
Woman	71 / 100 (71%)	34 / 50 (68%)	37 / 50 (74%)	
Man	28 / 100 (28%)	16 / 50 (32%)	12 / 50 (24%)	

Characteristic	**Overall**, N = 100[1]	**Testing**, N = 50[1]	**Training**, N = 50[1]	p-value[2]
Non-binary	1 / 100 (1.0%)	0 / 50 (0%)	1 / 50 (2.0%)	
Age	41.19 (12.55)	41.94 (13.40)	40.44 (11.72)	0.7
Baseline PHQ-9	12.47 (6.81)	11.56 (6.75)	13.38 (6.82)	0.2
Baseline GAD-7	12.89 (5.17)	12.80 (4.90)	12.98 (5.48)	0.8

[1]n / N (%); Mean (SD)

[2]Pearson's Chi-squared test; Fisher's exact test; Wilcoxon rank sum test

Facilitating therapeutic processes.

Pleasantries. Of the 50 clients sampled, 36% of participants sent messages that were considered pleasantries with their therapist. Overall, there were a total of 34 messages coded as "Pleasantries". Pleasantries indicate that the client is engaging in a kind, but not necessarily consequential way with their therapist, and can indicate rapport building and simple relationship maintenance. Some examples of messages that would be considered pleasantries are "Much appreciated, I will definitely reach out if I need to clarify anything. Thanks" and "Yeah, I can't wait. I haven't really been anywhere other than in the city during this whole quarantine, so definitely need a return to nature."

Response to Q from therapist (Neutral). Fourteen percent of participants sent messages that were coded as neutral responses to questions, with a total of 9 messages in this category. These messages indicate that the client is, at a minimum, engaged and willing to respond to inquiries from the therapist. Some examples of messages in this category are "Haven't tried. Going to this afternoon," and "Hello [FIRST_NAME]-- we are probably staying here till [DATE] or [DATE], I'm hoping for sooner."

Discussing positive aspects of the intervention. Twenty six percent of participants sent messages that were coded as discussing positive aspects of the intervention, with a total of 19 messages in this category. These messages indicate that the client is enjoying at least some part of the therapy intervention and communicating it with their therapist. Some examples of messages in this category are "Yes, I'm enjoying the program so far. I do think it's the right methodology for me, having to do the daily practices is definitely helping to build more of a routine into my day," and "The program has been helping me distract from stresses in life! I want to continue it."

Therapy interaction. Ten percent of clients sent messages that were coded as therapy interactions, with a total of 12 messages in this category. These messages indicate that the therapist and client are engaging in some sort of therapeutic intervention over chat. This is different than discussing situations and feelings (see below), in that it is not just the client describing their state, but working to understand this and how it is impacting their day to day lives. Some examples of messages in this category are "I have definitely noticed that I am on autopilot frequently, especially while I am home. It made me a little sad to realize how much I've been missing and how unengaged with my own life I have been. It I think that was a good realization for me to have so that I am more aware and work can work to stop being on autopilot," and "Not until something pulls me back to the moment. Then I realize I spent a lot of time doing nothing. I can go very deep into myself. I did this as a defense mechanism growing up. My father was a drug addict and alcoholic. There was a lot of trauma witnessing his self destructive behavior."

Discussing situations and feelings. Twenty two percent of clients sent messages that were coded as discussing situations and feelings, with a total of 22 messages in this category. These messages indicate that a client is describing a situation occurring in their life, often with feelings about the situation described in the message. Some examples of messages in this category are “The last couple of days have been rough for me. One of my friends has been blaming me for her suicidal thoughts and I had to block her. She still finds way to message me through my friends and social media. And then I got a warning at work for something I didn’t do. So my mind is absolutely spinning right now,” and “Lol broke my toe while walking in my house tonight. It's making me more stressed.”

Conceptual questions. Four percent of participants sent messages that were coded as conceptual questions, with a total of 2 messages in this category. These messages indicate that a client is asking questions about the intervention to better understand the intervention on a conceptual level. The two messages for this category were "I am still confused by HVR. So you want a higher number or your heart rate to be more variable instead of less variable? Seems opposite.” and “Not really getting my arms around the resonance concept. How is it supposed to feel? Not clear on what synching heart rate & breathing means or should feel like. Does it develop into a unconscious process over time with the practice? Any insights?”

Logistics questions. Ten percent of participants sent messages that were coded as logistics questions, with a total of 6 messages in this category. These messages indicate that clients need help solving a logistical issue in order to engage with the app and the intervention. Some examples of messages in this category are “Are we supposed to take our HRVB each day? I didn’t see where to log it,” and “I ended up not doing the HRV resting measure the moment I got up because I just wanted to make sure it worked. So should I redo it?”

Asking advice. No messages were sent that were categorized under this code.

Disrupting therapeutic processes.

Discussing difficulties with the intervention practices. Four percent of participants sent messages that were coded as discussing difficulties with the intervention, with a total of 2 messages in this category. These messages indicate some level of dissatisfaction with the intervention, specifically (as opposed to the app or some other difficulty in their life). Some examples of messages in this category are “I m dealing with some issues surrounding my breakup. So maybe processing that in addition to trying a new exercise is causing the added stress /anxiety” and “The other day I had a difficult time describing the emotion in the text box. I wanted to just say sad, but that’s also a feeling I didn’t feel all day. I was able change my mood so I didn’t want to put one feeling in there when that didn’t display the whole day.”

Tech difficulties. Eighteen percent of participants sent messages that indicated some level of technical difficulties, with a total of 13 messages in this category. These messages indicate some level of dissatisfaction or issues surrounding technology usage and app navigation. Some examples of messages in this category are “I thought I could practice hrv at any time by going to the program menu. I see the guided practice, but not the unguided practice,” and “I’ll try again to figure out app again tomorrow.”

Indicating busy-ness. Sixteen percent of participants sent messages that indicated that they were busy, with a total of 9 messages in this category. These messages indicate that they are having difficulties engaging with the intervention because of how busy they are in their daily lives. Some examples of messages in this category are “I worked last night 12 hours and slept 4 hours and work needs extra help so I am heading in to help so won’t be able to do the HRVB tonight. Hope u have great weekend,” and “Hey, sorry about my busy day yesterday. I'm on top of it today and I'm gonna be carving out time coming up more frequently.”

Concerns about the program. Two percent of participants sent messages that indicated that they had concerns about the program, with a total of 1 message in this category. These messages indicate that the clients feel negatively towards the program in some way. The one message in this category is "please call me I have concerns about the counseling and the program. If you can call me now I would be ever so grateful."

Requesting help. Four percent of participants sent messages that indicated that they needed help, with a total of 2 messages in this category. These messages indicate that clients were hoping to receive some sort of support from their therapist that was not related to the therapeutic interventions. The two examples of messages in this category are "No I'm ok but, you were going to send me some information. I didn't get anything from you?" and "[FIRST_NAME] please give me a call when you can thanks."

Neutral processes.

Other. Four percent of participants sent messages in the "other" category, with a total of 2 messages in this category. The messages were mostly correcting typos. Some examples of messages in this category are "*beat brusher" and "Idk."

Statistical Analyses

Participants.

For these analyses, there were 11,210 messages sent between therapist and client and there were a total of 762 participants nested within 15 therapists. See Table 18 for descriptive statistics for the sample used in these analyses. There was a statistically significant difference in age between those who completed and those who dropped out, with those who dropped out being, on average, younger than those who completed.

Table 18.
Summary of descriptive statistics for the participants used in quantitative analyses.

Characteristic	**Overall**, N = 762[1]	**Completer**, N = 564[1]	**Drop-out**, N = 198[1]	**p-value**[2]
Age	40.82 (11.60)	42.21 (11.78)	36.86 (10.09)	<0.001
Gender				0.8

Characteristic	**Overall**, N = 762[1]	**Completer**, N = 564[1]	**Drop-out**, N = 198[1]	**p-value**[2]
Woman	582 / 762 (76%)	427 / 564 (76%)	155 / 198 (78%)	
Man	178 / 762 (23%)	135 / 564 (24%)	43 / 198 (22%)	
Non-binary	2 / 762 (0.3%)	2 / 564 (0.4%)	0 / 198 (0%)	
Baseline PHQ-9	11.48 (6.05)	11.30 (5.98)	12.00 (6.23)	0.2
Baseline GAD-7	11.57 (4.96)	11.41 (4.94)	12.03 (5.01)	0.11

[1]Mean (SD); n / N (%)

[2]Wilcoxon rank sum test; Fisher's exact test

Models.

For the first hypothesis, a poisson generalized linear model was conducted. The results indicated that there was a significant negative relationship between therapist message length and number of texts received from their clients (see Table 19, see Figure 14). This indicates that, as therapist message length increases, there is a decrease in the amount of messages that a client sends.

Table 19.
Results of a Poisson generalized linear model examining the relationship between therapist message length and client message count.

	Client Message Count				
Predictors	*Incidence Rate Ratios*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	13.10	13.09	10.31 – 16.65	10.31 – 16.64	**<0.001**
Therapist Message Length (scaled)	0.92	0.92	0.90 – 0.95	0.90 – 0.95	**<0.001**
Random Effects					
σ^2	0.07				
$\tau_{00\ \text{therapist}}$	0.22				
ICC	0.75				

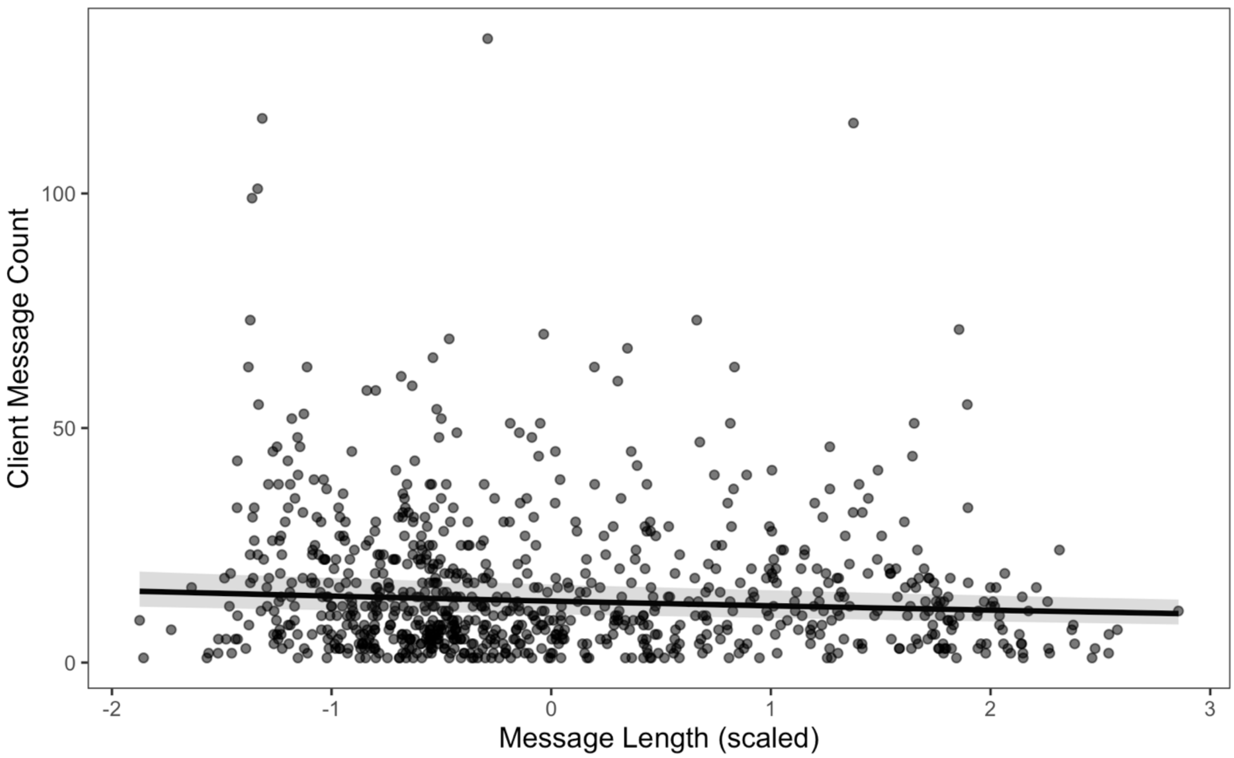

Figure 14.
Scatter plot of a Poisson generalized linear model examining the relationship between therapist message length and client message count.

For the second hypothesis, a binomial generalized linear mode was used to determine if there was a decreased likelihood of clients responding to questions within longer messages. Results indicate that clients were less likely to respond to a question if the message was longer (see Table 20, Figure 15).

Table 20.
Results of a binomial generalized linear model examining the relationship between therapist message length and if a client responded to a question.

	Client Response to Question				
Predictors	*Odds Ratios*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	1.00	0.95	0.76 – 1.32	0.72 – 1.26	0.986
Therapist Message Length (scaled)	0.81	0.80	0.77 – 0.85	0.76 – 0.84	**<0.001**

Random Effects

σ^2	3.29
$\tau_{00\ \text{userID:therapist}}$	0.68
$\tau_{00\ \text{therapist}}$	0.25
ICC	0.22

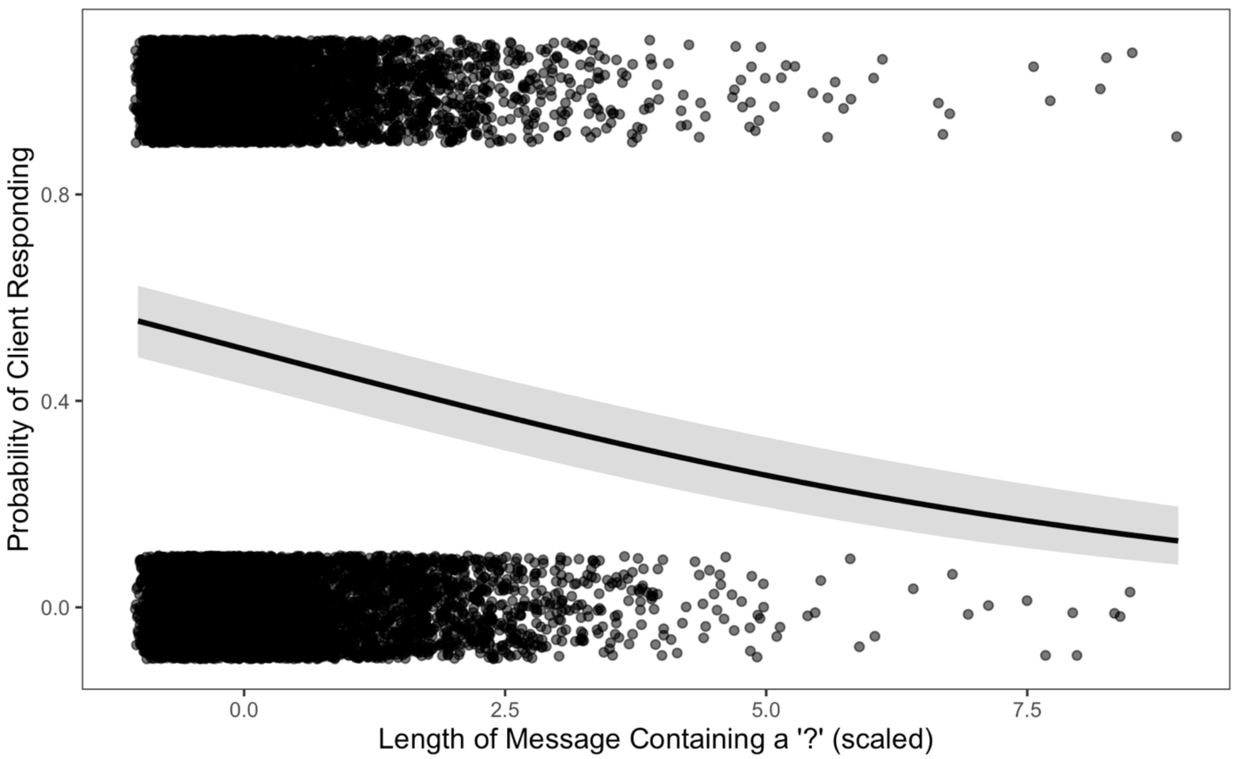

Figure 15.
Figure depicting the probability of a client responding to a question depending on the length of the message.

Machine Learning

Table 21 shows the statistics associated with the sample of participants used for these analyses. There was a significant difference in age between completers and drop-out clients, with completers being, on average, older. No other significant differences were observed between completers and drop-outs. The topic modeling procedure for the therapist's messages calculated 47 separate topics within the documents (messages) from therapists. Of the 47 topics, 14 topics were significantly related to completion of the MHP, with 7 being negatively related to completion and 7 being positively related to completion. Of note, the two strongest topics negatively related to completion have themes of calling the client and apologizing for difficulties within the app (see Table 22 Example Messages column). The two topics that were positively related to completion were related to checking in and offering support.

The topic modeling procedure for client's messages calculated 32 separate topics within the documents from clients. Of the 32 topics, 20 were related to completion of the MHP, with 12 being negatively related to completion (see Table 23). Themes for the strongest topics negatively associated with completion were around quitting and heart rate variability monitors. The two topics that were positively related to completion were related to reaching out and good weekend.

Table 21.
Summary of descriptive statistics for the participants used in Machine Learning analyses.

Characteristic	**Overall**, N = 762[1]	**completer**, N = 564[1]	**drop-out**, N = 198[1]	**p-value**[2]
Age	40.82 (11.60)	42.21 (11.78)	36.86 (10.09)	<0.001
Gender				0.8
Woman	582 / 762 (76%)	427 / 564 (76%)	155 / 198 (78%)	
Man	178 / 762 (23%)	135 / 564 (24%)	43 / 198 (22%)	
Non-binary	2 / 762 (0.3%)	2 / 564 (0.4%)	0 / 198 (0%)	
Baseline PHQ-9	11.48 (6.05)	11.30 (5.98)	12.00 (6.23)	0.2
Baseline GAD-7	11.57 (4.96)	11.41 (4.94)	12.03 (5.01)	0.11

[1]Mean (SD); n / N (%)

[2]Wilcoxon rank sum test; Fisher's exact test

Table 22.
Table of the results for each topic from the Structural Topic Modeling related to completion within the therapist text messages.

Topic	effect	p value	Example Messages	Topic words
15	-0.01555	< .001	I appreciate you sharing your concerns about the program. It may not address all depression and behavior change [DATE], so we can talk more about expectations. We have developed this program based on research about the effectiveness of these on - line therapies. So, there is good evidence both with our program and in the research that these tools can help when practiced consistently! Everyone is unique and that's why I am here, to collaborate with you in making this program the best it can be for your needs. :)	call, hang, inform, full, help, can, journal
2	-0.01329	< .001	Hi [NAME], I'm sorry you're encountering some challenges. A few thoughts: 1. Be sure you are on the " [DATE] " screen. You can toggle to that screen at the very bottom of the app where you will see icons that say " [DATE] ", " group ", " program ", " notifications ". From the [DATE] Screen you will see two practices for [DATE]. You can click the green circle icon to get to these practices.	apolog, bottom, end, motiv, tab, show, screen
22	-0.01116	< .001		monkey, simpli, wander, hand, notic, distract, common,
29	-0.01015	< .001		rhythm, biofeedback, rate, pressur, differ, locat, true
43	-0.00896	< .001		updat, resolv, thank, much, name, head, possibl
9	-0.00895	< .001		mood, diari, low, impact, struggl, worri, believ
4	-0.00850	< .001		awesom, work, great, wonder
14	0.00681	< .001		stick, tough, howev, time, system, heart, case,
10	0.00773	.003		hijack, 'll, stop, develop, distanc, skill, cant,
31	0.00947	< .001		pick, count, everi, exampl, congratul, rememb, practic,
20	0.00956	.003		download, click, meru, phone, pair, correct, devic,
17	0.01585	.015		log, orient, offici, wish, chanc, login, welcom,

Topic	effect	p value	Example Messages	Topic words
7	0.01633	< .001	Hi [NAME], thank you for letting me know. Let me know how the call with the ENT goes. It sounds like you have been on top of this as much as possible. Hang in there and let the doctor's do their job!	let, know, flexibl, goe, there, cours, support,
21	0.02856	< .001	Hi [NAME], I just wanted to check - in and see how the program is going for you so far? Any troubles or issues I can help support you with?	check, anyth, want, dont, recent, miss, just,

Table 23.
Table of the results for each topic from the Structural Topic Modeling related to completion within the therapist text messages.

Topic	effect	p value	example_messages	topic_words
13	-0.02089	< .001	Unfortunately there isn't. This job is non - stop and not the best for someone like me. It wasn't like this before Covid and shutdown. It has changed a lot. Once the 2nd Supervisor is trained I'm hoping things will be better.	text, option, quit, session, cant, give, say,
11	-0.01743	< .001	Hey. My monitor won't keep or take a charge. I'm unable to do the biofeedback.	

Topic	effect	p value	example_messages	topic_words
7	-0.00666	.007		result, wait, hear, can, see, didnt, love,
2	-0.00543	.023		awesom, check, name, tip, feedback, okay, ,
26	-0.00471	< .001		get, wrong, day
14	0.00380	.030		set, min, alreadi, notif, peopl, focus, tell,
9	0.00696	.012		husband, hard, meet, famili, difficult, year, time,
5	0.00939	.012		want, els, let, know, sure, anyth, just,
18	0.00998	.002		sorri, place, busi, super, emot, made, got,
20	0.01035	.018		negat, self, drive, bring, lost, posit, live,
23	0.01291	.021		phone, receiv, connect, devic, app, troubl, hrv,
6	0.01593	< .001	So far so good, thanks for checking and you have a good weekend too!	expect, good, hope, weekend, far, share,
15	0.05431	< .001	Hi [NAME]! Thanks for reaching out. Yes, I will review the material [DATE]. [DATE] was a long day but [DATE] is a little lighter.	email, much, yes, perfect, date, far, jen,

Study 2

Method

Measures

Completion rates. Completion rates were defined as completing at least half of the instructional weekly videos during the course of the program.

Engagement variables. All behavioral variables were taken from the first week of participating in the MHP. Active Days is defined as the amount of days within the first week interacting with the material on the MHP platform. Chat messages sent is the number of messages that were sent by participants within the first week. Finally, as there are modules within the platform to meditate and engage in Heart Rate Variability Biofeedback (HRVB), we were able to create a variable capturing the total time spent meditating during the first week and total time spent in HRVB during the first week.

Self-Report Variables. Depression was measured using the PHQ-9, and Anxiety was measured using the GAD-7 (see previous section). Motivation was measured by asking participants at the beginning of the MHP on a scale of 1-10 "question?"

Data analysis

In order to determine which initial indicator was the best predictor of completion, we ran a series of logistic regressions for each predictor. Next, the Active Days, Minutes Meditating, HRVB Minutes, and Messages sent were grouped into a single "engagement" logistic regression representing the variables that are indicative of objective behavioral measures within the first week of the program. Another "self-report" logistic regression was created using the PHQ-9 and GAD-7 baseline scores. These two models (engagement and self-report models) were then compared to determine which better explained the variance between completers and non-completers using BIC (which accounts for differences in the amount of predictors across models). Finally, all of the predictors were added to a single logistic regression in order to determine which variables were still significantly related to completion. Standardized betas were calculated in order to make comparisons between the strength of the predictors. Additionally, model checks were utilized in order to determine variable inflation factor (VIF) for each predictor to ensure that there were no issues with multicollinearity (VIF <5). All models included age and gender as covariates.

Results

Of those that participated in the MHP, there were 900 (72%) participants who completed the program and 356 (28%) participants who did not complete. Separate Wilcoxon tests revealed that completers were higher, on average, in age, days active in the app, time spent meditating and in HRVB, and sent more messages. Completers were likely to start out with a lower PHQ-9 score. See table 24 for descriptive statistics.

Table 24.
Summary of descriptive statistics for the participants.

Characteristic	**Overall**, N = 1,256[1]	**drop-out**, N = 356[1]	**completer**, N = 900[1]	**p-value**[2]
Age	38.66 (10.90)	36.23 (10.04)	39.62 (11.08)	<0.001
Gender				0.7
Woman	999 / 1,256 (80%)	282 / 356 (79%)	717 / 900 (80%)	
Man	241 / 1,256 (19%)	71 / 356 (20%)	170 / 900 (19%)	
Non-binary	16 / 1,256 (1.3%)	3 / 356 (0.8%)	13 / 900 (1.4%)	
Active Days	4.54 (1.78)	3.36 (1.76)	5.00 (1.56)	<0.001
Minutes Meditating	11.19 (13.49)	6.04 (7.41)	13.23 (14.76)	<0.001
HRVB Minutes	17.92 (12.87)	10.97 (9.99)	20.67 (12.85)	<0.001
Messages Sent	3.54 (3.46)	2.83 (2.98)	3.82 (3.59)	<0.001
Baseline GAD-7	11.87 (4.98)	12.16 (5.29)	11.76 (4.85)	0.12
Baseline PHQ-9	12.11 (6.03)	12.95 (6.26)	11.78 (5.91)	0.003

[1]Mean (SD); n / N (%)

[2]Wilcoxon rank sum test; Fisher's exact test

For the separate logistic models, days spent using the MHP during the first week was significantly predictive of completer status ($b = 0.55$, 95% CI $[0.47, 0.63]$, $z = 13.29$, $p < .001$; Fig. 16A). Additionally, time spent meditating during the first week was significantly predictive of completer status ($b = 0.07$, 95% CI $[0.05, 0.09]$, $z = 7.73$, $p < .001$; Fig. 16B), as was amount of activity in chat with the therapist ($b = 0.10$, 95% CI $[0.06, 0.15]$, $z = 4.28$, $p < .001$; Fig. 16C). Finally, time spent in HRVB practice significantly predicted completer status ($b = 0.07$, 95% CI $[0.06, 0.08]$, $z = 11.16$, $p < .001$; Fig. 16D). From the mental health measures, baseline PHQ-9 scores were significantly related to completer status ($b = -0.03$, 95% CI $[-0.05, -0.01]$, $z = -3.07$, $p = .002$; Fig. 16E), but baseline GAD-7 scores were not ($b = -0.01$, 95% CI $[-0.04, 0.01]$, $z = -0.95$, $p = .342$; Fig. 16F).

When combining variables into two separate models (i.e., engagement model vs. self-reported model), the objective reports model (4 predictors of interest, BIC = 1,293.97) explained significantly more of the variance than the self-reported model (2 predictors of interest, BIC = 1,503.65). Finally, all of the predictors were added into a single model (along with the previously mentioned covariates) to see which variables remained a significant predictor of completer status. Table 25 shows the standardized betas for each predictor, allowing for comparison between the predictors. All predictors were significantly associated with the completion, with the exception of baseline GAD-7. Additionally, Active Days had the highest standardized beta, followed by Minutes Meditating. Model checks showed that there were no issues with multicollinearity (as measured by variable inflation factor).

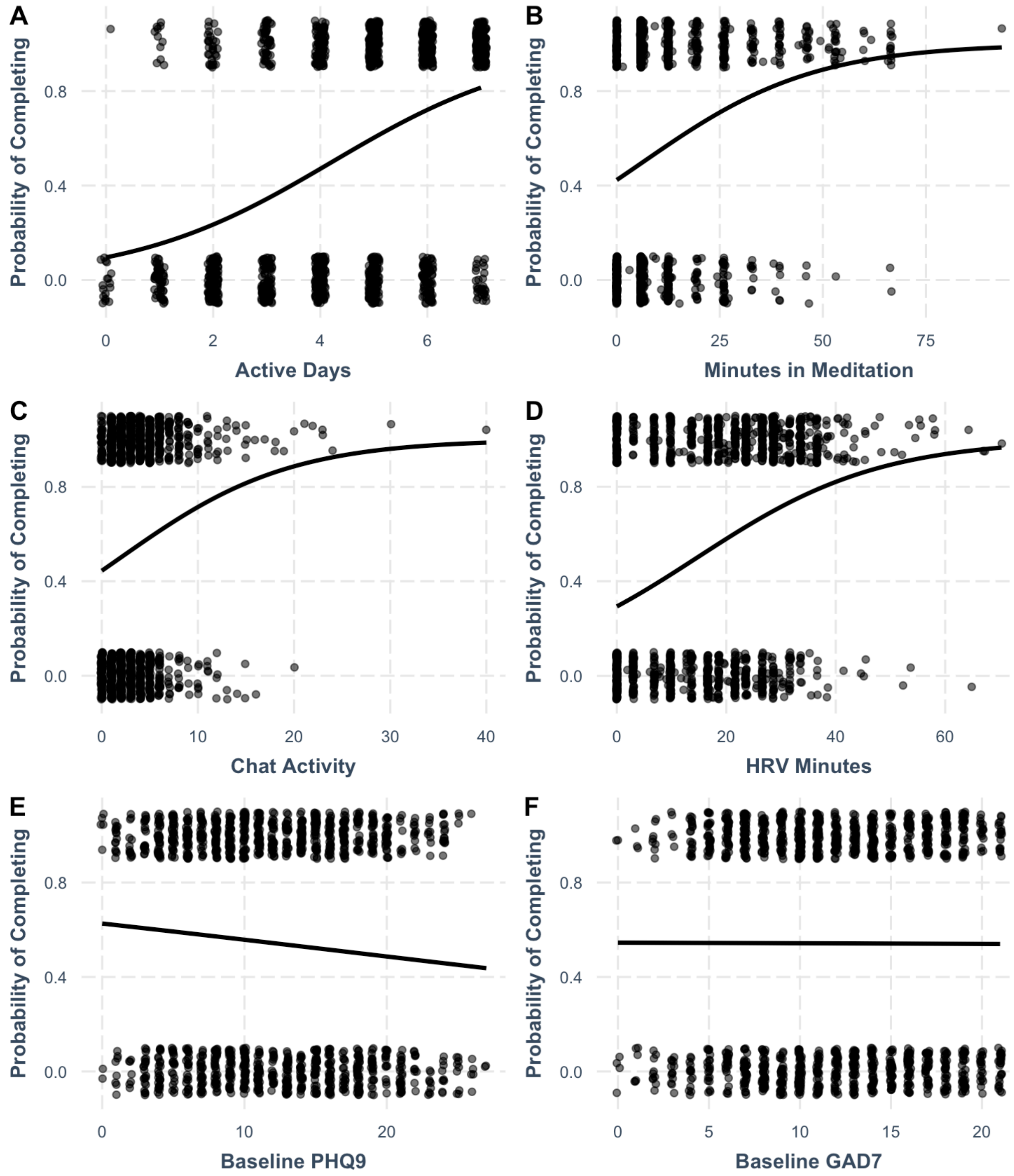

Figure 16.
Results of 6 separate logistic models with each separate predictor variable predicting the likelihood of completion.

Table 25.
Results from a logistic regression with all the predictors included and standardized betas, with completion as the predicted variable.

Predictor	Estimate	Std. Error	z value	*p* value
(Intercept)	1.20	0.08	15.13	$p < 0.001$
Active Days	0.66	0.11	6.04	$p < 0.001$
Minutes Meditating	0.37	0.12	3.16	$p < 0.001$
Messages Sent	0.19	0.09	2.25	0.02
HRVB Minutes	0.25	0.11	2.25	0.02
Baseline PHQ-9	-0.16	0.08	-2.03	0.04
Baseline GAD-7	0.05	0.08	0.58	0.56

Discussion

In order to determine how clients were engaging with the MHP, I employed several techniques to measure engagement for Study 1. The qualitative analyses helped provide a better understanding of the ways in which clients were messaging with therapists within the first week of the program. I discovered that the plurality of messages that were sent from individuals were pleasantries. This makes sense, given the nascent nature of the relationship. Many of the messages, overall were positive, with the most frequent being the discussion of positive aspects of the intervention, discussing situations and feelings with their therapist, and responding to prompts/questions from the therapists. Even if some of the content of these messages are negative (e.g., discussing negative feelings and situations), these are helpful for facilitating connection with the therapists, which is a positive predictor for change within a therapeutic context (D’Alfonso et al., 2020; Dixon et al., 2016). Additionally, there were messages that fell under the category of “disrupting the therapeutic process.” Many people indicated that they had issues with technology, difficulties engaging with the therapeutic practices, or that they were very busy and having a difficult time keeping up with the content of the program, given their schedule. All of this indicates that clients are sharing a variety of sentiments and topics with their therapists, which is, overall, beneficial for. Understanding what clients are enjoying from the program, how they are benefiting from the therapeutic process, and some of the difficulties they are encountering within the program.

For the second portion of the study, statistical analyses were conducted to see which messages clients were responding to from their therapists. The first analyses demonstrated that therapists who send longer text messages receive fewer client messages than therapists that send shorter text messages. The second analyses demonstrated that, for messages that contained a question, the longer the message that contained the question, the less likely it was for a client to respond to that question. Overall, these results suggest that shorter messages are associated with increased responsiveness from clients.

For the third portion of study 1, I used a machine learning algorithm that detected topics within the text dataset for both therapist messages and client messages. Messages from therapists that were negatively related to completion rates were interestingly about therapists calling clients and apologizing for difficulties within the app. Of note, therapists in the MHP rarely call the client unless there are issues. In particular, clients are asked to schedule a call when they are non-responsive within the app. It is generally a positive sign that this is the topic that is most closely related to drop-out within the program, as this demonstrates that the algorithm is detecting a signal from the noisy text data. Inversely, the therapist's messages that were most related to completion were related to checking in and offering support, which suggests that therapists checking in with clients and offering support about how they are doing in their daily lives and within the program are related to completion. It should be pointed out that this does not necessarily mean this is causal in nature. It could be that clients that are more engaged with the program and their therapists *elicit* these types of messages from therapists, and these types of clients are, in turn, more likely to complete the program. For the themes related to dropout and completion in the client's messages, the messages are a little less clear. Themes related to drop-out were quitting and HRVB monitors. The theme of quitting makes sense within this context, but, with HRVB monitors, it could be that difficulties with these devices (and subsequent frustration) could be the predictive part of drop-out. Themes related to completion were "reaching out" and "good weekend." This could be indicative of clients having conversations and building rapport with their therapist, which in turn makes them more likely to complete the program, but any causality would need to be explored with future studies.

Finally for study 2, Active days was shown to have the strongest predictive power of all the predictors for dropout. Additionally, when examining the difference between behavioral engagement variables and self-report variables, engagement variables were better able to explain the variance in completion rates than self-report variables. This could be indicative that, if there is interest in determining who might be at risk of dropping out, utilizing behavioral variables would most likely be a better strategy for predicting drop-out than using self-report variables. This is consistent with the broader literature at large, as behavioral variables are often more closely related to outcomes than self-report variables (Lewandowski & Strohmetz, 2009).

There are several important implications for these two studies. Firstly, to my knowledge, this is the first study that has ever examined message data between therapist and client in a real-world consumer-available intervention. This is why it was important to examine the chat messages using multiple methodologies. The qualitative analyses provide insights into what is being discussed between therapist and client. The topics discovered offer a glimpse into the ways in which clients communicate the benefits of the program, difficulties they are having with the program, and other aspects of their lives that are relevant to their treatment. This also lends credence to the notion that the therapeutic relationship is helpful for some within the MHP, as they are utilizing the chat function to improve their experience within the program. Another implication from these studies confirms a long-standing understanding the field has had for a long time: behavioral measures tend to be better predictors of outcomes than self-report variables (Lewandowski & Strohmetz, 2009). This not only provides evidence that bolsters those previous findings, but also expands this into the digital mental health space. Future clinicians and researchers can potentially use this as a stepping-stone for building models that will be predictive of who might drop out of the intervention. This would be greatly beneficial, as this could be a first-warning system designed to help those at risk of not completing the intervention, meaning that they may not receive the full effect of the intervention.

There are several limitations to these studies, however. For the first study, the first major limitation is that each individual analysis only provides a small window into the data. Given the vast nature of the message data and the current technology, it is difficult to process 10,000+ messages computationally, let alone using existing qualitative methods. The qualitative analyses only examined messages from 50 participants (significantly fewer than the overall sample), meaning that there are certainly many topics that were not discovered in the explorative process in the qualitative analysis. Additionally, for the statistical portion of study 1, working with text data is, for lack of a better term, messy. For example, with the following methodology (discussed in Method section above):

> A binary variable was created that accounted for whether or not the message following the question from the therapist was from the client or the therapist at a later time. If the

following message was from the client, this was considered a "response" to the therapist's question. If the following message was from the therapist at a later time, this was considered "no response" to the therapist's question.

With this methodology, there is no way of assuredly *knowing* that the message from the client is a response to the question from the therapist. Certainly, there must be instances where the client is messaging about another topic entirely. But the assumption is that the majority of texts from the client are a response to these questions from the client, and that there are enough messages for a signal to be detected in spite of the occasional violation of this assumption.

Additionally, another limitation is using the first week worth of data for our analyses. This was done in order to use data from *only* within the first week to examine how new participants might be engaging with the therapist and the program as a whole, as well as begin testing which variables might be useful in predicting dropout, but it does limit our ability to make inferences about all of the data. For instance, participants could (and most likely do) change the way they interact with the app or have different types of conversations with their therapists after the first week. Based on these current analyses, these changes are not observed and unaccounted for. Given these limitations, it should also be noted that there are some strengths to these studies. These studies utilized a mixed-methods approach, which can be helpful for using different lenses with which to understand and interpret the data. This provides a more complete picture than if utilizing one method to examine the data. Additionally, given that the analyses often used over 700+ participants (with repeated measures in some instances), this provided more than enough power to detect even small relationships between the variables of interest.

Overall, these two studies add to our understanding of how individuals engage within a DMHI. Clients are willing to discuss with their therapists the benefits of the therapeutic intervention, as well as the pain points. Additionally, clients are able to engage in rapport building with the therapist by exchanging social niceties. Clients, however, do seem to have difficulties responding to longer messages from therapists and are not likely to respond to questions embedded within longer messages. Finally, when using variables obtained within the first week, behavioral engagement variables are more reliably a predictor of drop-out than self-reported mental health measures. More specifically, the number of days that a client is active within the app is the most predictive of whether a participant will complete the program or not. Overall, these findings can be used to better inform the development and implementation of DMHIs in the future to come.

Chapter 4

Introduction

Over the past decade, there has been a substantial increase in reported mental health difficulties (Goodwin et al., 2020); however, due to high costs, poor psychotherapist-to-patient ratios, shortage of psychotherapists in specific geographical regions, and a myriad of other structural barriers, many patients find it difficult to access mental health services (Alang, 2015; Alegría et al., 2018; Fumarco et al., 2020; Kirby et al., 2019). Interestingly, recent advances in mobile health (mHealth) technology, combined with the rapid adoption of smartphone technologies, have opened up a new avenue to address rising mental health needs. Digital mental health treatments provide an opportunity to decrease barriers to care in order to more seamlessly integrate mental health care into patient's lives (Schueller et al., 2019). Not only does digital mental health provide access to those who need it, but it also provides affordable and scalable interventions that can be tailored and adapted to the individual. Additionally, with the recent Coronavirus pandemic, digital mental health treatments are in higher demand than ever before. However, digital mental health interventions face the challenge of translating traditionally in-person interventions in a way that can be readily accessed in one's daily life. One intervention, in particular, that shows promise for both effectiveness and easy adaptation is the use of heart rate variability biofeedback (HRVB), which involves providing participants with a wearable device that provides real-time visualization of beat-to-beat heart rate intervals during slow, paced breathing.

HRVB has recently been added as an adjunctive and supportive component to mental and physical health interventions (P. Lehrer & Gevirtz, 2014; P. Lehrer et al., 2020). HRVB practice may provide one avenue for volitionally upregulating parasympathetic control, leading to possible cognitive and affective benefits (P. Lehrer & Gevirtz, 2014). However, HRVB treatment typically involves careful calibration of one's optimal breathing pace, in order to maximize the efficacy, which makes it challenging to offer at scale and at low cost. Digital mental health interventions are poised to address the significant challenge of meeting the need for mental health services at scale, and might benefit from the addition of HRVB. However, it is unclear whether HRVB can contribute to symptom reduction in a naturalistic environment, using a predetermined breathing pace of 6 breaths per minute, and given the expected degree of environmental "noise" in HRV, when measured under real-world conditions.

HRV and Mental Health

HRV represents the temporal variation in beat-to-beat intervals between consecutive heart beats and is thought to be controlled by vagally mediated higher-order inhibitory cortico-subcortical neural circuitry spanning the medial and orbitofrontal aspects of the prefrontal cortex, anterior insula, cingulate cortex, central nucleus of the amygdala, and the brainstem (Kemp & Quintana, 2013; Thayer et al., 2009). Many of these same brain regions have been implicated in emotion regulation and cognitive control, which may explain why higher resting HRV (i.e., greater variability in heart rate fluctuations) is associated with greater emotion regulation and psychological flexibility (Kemp & Quintana, 2013; Mather & Thayer, 2018). In contrast, lower HRV has been found in a host of psychiatric disorders, including depression (Kemp et al., 2010), anxiety (Chalmers et al., 2014), Bipolar disorder (Henry et al., 2010; Lee et al., 2012), Attention Deficit Hyperactivity disorder (Buchhorn et al., 2012), Schizophrenia (Berger et al., 2010), Alcohol Use disorders (Quintana et al., 2013), and Conduct disorder (Beauchaine et al., 2007). Furthermore, severity of psychiatric disorder appears to be associated with lower HRV, such that there is a negative correlation between depression severity and lower HRV (Kemp et al., 2010).

Additionally, even in non-clinical populations, lower HRV has been associated with higher state and trait anxiety scores, as well as depressive scores (Shinba et al., 2008). Owing to the links between HRV and mental health disorders, researchers have started to examine whether improvements in HRV could have ameliorative effects on mental health symptoms.

HRV can be modified by various biopsychosocial factors, including environmental, psychological, and physical stress (Shaffer & Ginsberg, 2017), and is thought to be regulated by cognitive and affective neural regions of the brain (Mather & Thayer, 2018; Shaffer & Ginsberg, 2017; Shaffer et al., 2014; Thayer et al., 2009). In general, higher HRV has been associated with positive health outcomes and is thought to indicate a greater ability to flexibly adjust to environmental demands, while lower HRV is associated with deleterious mental health outcomes as well as subsequent morbidity and mortality (Kemp & Quintana, 2013). Therefore, HRV may serve as a putative biological mechanism whereby stressors exert effects on individuals as explained by the stress diathesis model of health (Kemp et al., 2017). As such, HRV has been conceptualized as a transdiagnostic biomarker for mental health outcomes (Beauchaine & Thayer, 2015; Kemp & Quintana, 2013).

HRVB as an Adjunctive Treatment for Depression.

Recently, there has been a push to provide HRVB to a wider range of individuals. As such, more cost effective Bluetooth heart rate sensors have entered the market (e.g., Kyto HRV Monitor), making HRVB much more accessible. During HRVB practice, patients are shown real-time heart rate data and prompted to breathe at a prescribed rate corresponding to the resonance frequency of the baroreflex system, which is directly stimulated in HRVB (E. G. Vaschillo et al., 2006). The large-amplitude oscillations in heart rate achieved by this procedure exercise and strengthen the baroreflexes (P. M. Lehrer et al., 2003), which directly modulate swings in blood pressure and, indirectly, other homeostatic processes throughout the body (Eckberg & Sleight, 1992). There are direct anatomical projections from this system between brain centers that generate anxiety and depression (the insula and amygdala) and those that control it (prefrontal cortex) (Mather & Thayer, 2018). Connectivity between these centers is increased by regular practice of HRVB (Nashiro et al., 2022), and HRVB has actually been found to increase the mass of brain tissues in the prefrontal cortex (Yoo et al., 2022). This may explain the moderate to large effect sizes found in meta analyses for HRVB effects on anxiety, depression, and other negative emotional states (Fernández-Alvarez et al., 2022; Goessl et al., 2017; P. Lehrer et al., 2020). For an in-depth HRVB protocol and further explanation, see Lehrer et al. (2013). However, when translated into a digital therapeutic context, it would be difficult to conduct a sufficiently thorough calibration to determine a participant's ideal resonance frequency. Hence, one alternative would be to test whether having patients breathe at a pace that would place the majority of the population roughly within the zone of resonance (i.e., 6 breaths per minute) would still provide clinically significant benefits, although there is evidence that breathing slightly faster than resonance frequency (Steffen et al., 2017) has smaller effects on anxiety, even if still significant, and that breathing closer to resonance also is also associated with increased blood pressure control when compared with a 6 bpm breathing exercise (Lin et al., 2014).

However, while there have been studies that have shown that HRVB is an effective treatment component for reducing mental health symptoms, it is also important to determine which aspects of physiology are changing in conjunction with mental health symptoms. In particular, it would be useful for the field to have potential biomarkers that could be used to identify physiological mechanisms of change related to reductions in anxiety and depressive symptoms. One potential biomarker that could prove useful is the Power Spectral Density (PSD) within the low frequency (LF) band, as this corresponds to increasing baroreflex gains. Previous studies have found that, when an individual breathes at their resonance frequency, the amplitude of their heart rate oscillates at a higher frequency (Sakakibara et al., 2020). Another study found that even breathing at frequencies near resonance frequency produced a larger LF amplitude than breathing at further away from resonance frequency (E. Vaschillo et al., 2002). One study, using a non-clinical sample, found that increasing power in the LF range in a HRVB training group (compared to control group) was related to reductions in depression and anxiety symptoms (Sutarto et al., 2012). Because increasing spectral distributions within the LF amplitude are related to resonance effects involving both respiratory sinus arrhythmia (RSA) and baroreflex gain, this could serve as a useful physiological marker related to changes in mental health symptoms.

Current Study

While initial studies indicate that HRVB may be a promising potential adjunctive treatment to gold-standard interventions, more research is required to examine these effects in sufficiently powered studies, especially within the context of the nascent digital mental health treatments. In the current study, we sought to examine whether HRVB practice, defined as engaging in 6 breaths per minute (bpm) resonance frequency breathing as part of the Meru Health Program (MHP), an evidence-based digital mental health intervention, was associated with significant reductions in anxiety and depressive symptoms. Specifically, we investigated whether a physiological indicator of breathing within the resonance zone (PSD around the periodicity of 6 bpm) was related to decreases in depression and anxiety symptoms. In particular, we wanted to test whether both 1) initial differences in the PSD of 6 bpm resonance zone breathing at an individual's very first HRVB session at the start of treatment , and 2) the improvements in PSD of 6 bpm resonance zone breathing across multiple HRVB sessions throughout treatment were related to reductions in depression and anxiety symptoms during a digital mental health intervention. We hypothesized that positive physiological changes (e.g., increases in PSD around 6 bpm) throughout the program would be associated with significant decreases in both anxiety and depressive symptoms from baseline to follow-up. We did not, however, have a priori hypothesis for whether the initial ability to breathe in synchrony around the periodicity of 6 bpm, measured by PSD at 6 bpm, would be related to depression or anxiety symptom reduction.

Methods

Participants and Recruitment.

The MHP has been described in detail in prior publications (Economides et al., 2020, 2019; Goldin et al., 2019). Briefly, the MHP is a 12-week digital mental health intervention (originally developed as an 8-week program) with evidence-based components delivered asynchronously via a smartphone app. The program incorporates a continuous care model that includes frequent interaction with a dedicated, licensed clinical therapist and as-needed consultations with medical doctors, including psychiatrists. Patients presented to the MHP via several different avenues. Most patients are referred by healthcare providers or employee assistance programs. Inclusion/exclusion criteria of the MHP require patients to have at least mild levels of depression, anxiety, or burnout, own a smartphone, and not have an active substance use disorder, severe active suicidal ideation with a specific plan, severe self-harm, or a history of Bipolar Disorder or psychosis. All patients enrolled signed informed consent to participate, have their data collected, and allow for de-identified data to be used for research purposes. Data collected as part of care, including engagement data, HRVB data, and depression and anxiety outcomes data, are stored in Health Insurance Portability and Accountability Act (HIPAA) compliant electronic medical records that includes protected health information. All measures examined in this study were collected in the MHP app before, during, and at end-of-treatment. Institutional review board exemption for this analysis was granted by the Pearl Institutional Review Board for analyses of previously collected and de-identified data.

At the onset of the MHP, each patient is sent a heart rate sensor (see below for more details). The sensor is included with the treatment and does not require an additional purchase from the patient. In addition, patients receive a brief written and video introduction to HRV and resonance breathing, including how to use the device. The HRVB component of the program, referred to as "resonance breathing exercises," starts at a duration of 5 minutes and gradually increases up to a maximum of 20 minutes per day, with patients directed to adjust the duration in 5-minute increments accordingly (though patients are allowed to go shorter or longer if desired). Patients are encouraged to engage in HRVB training roughly 2-3 times a week.

Patients were included in analyses if they had participated in the MHP after December 02, 2018 and before November 11, 2020. Our initial sample consisted of 588 patients; however, 201 patients did not have 6 or more HRVB sessions, which was necessary for estimating a slope for an individual (see Statistical Analysis for more information). After excluding the 201 patients with less than 6 HRVB sessions (see Statistical Analysis), the analytic sample was left with 387 patients (Mean age = 39.7, SD = 9.6, Range: 20 – 67; 79.3% females), who collectively had 16,109 HRVB sessions. While this exclusion rate is higher than other–studies that have examined HRVB training, this is a field study that emphasized HRVB as an adjunct to treatment and not the main treatment itself.

There were significant differences between those who were included and those who were excluded based on this criterion, with included patients having more total active days in the MHP ($W = 12{,}094.5$, $p < 0.001$), being younger ($W = 34{,}676$, $p = 0.024$), and being more likely to complete the MHP ($\chi^2(1) = 183.28$, $p < .001$) than excluded patients. No other significant differences were observed.

There were also significant differences between individuals in the two different versions of the MHP (8-week versus 12-week; see below section for more information regarding program weeks) for total HRVB minutes ($W = 6{,}697.00$, $p < .001$), baseline PHQ-9 scores ($W = 6{,}420.00$, $p < .001$), program location ($\chi^2(1) = 275.03$, p < .001), and gender ($\chi^2(1) = 4.35$, $p = .037$).

Assessment & Treatment Procedures.

We examined existing baseline and end-of-treatment clinical outcome data collected from patients treated with the MHP, as well as daily HRVB data. Traditionally, when HRVB is used, an individual is assessed for their personalized resonance zone. However, when translated into a digital therapeutic context, it would be difficult to conduct a sufficiently thorough calibration to determine a participant's ideal resonance frequency. Hence, we currently instruct patients to breathe at a pace that would place the majority of the population roughly within the zone of resonance (roughly 6 bpm), though they are instructed that if this is uncomfortable they can try and find a breathing pace that feels more natural. During each HRVB session, patients are guided by a visual pacer that expands during inhalation for 4 seconds and contracts during exhalation for 6 seconds, achieving resonance breathing at a rate of approximately 6 breaths per minute. The visual pacer is supplemented with recorded breaths that matched the rate of the visual pacer. Below the pacer, patients are shown a real-time visual trace of their heart rate which is green during periods of high resonance and amber during low resonance. At the end of the practice patients are shown summary feedback detailing the session duration and time spent in high and low resonance.

Measures.

Depressive and Anxiety Symptoms. Depressive symptoms were measured at baseline and every 2 weeks through the program's end by the Patient Health Questionnaire-9 (PHQ-9), a widely used instrument used to screen for depression (Kroenke et al., 2001). The PHQ-9 consists of a list of nine depressive symptoms with response options ranging from 0 (not at all) to 3 (nearly every day). The PHQ-9 has excellent internal consistency (Cronbach's α of 0.89 in primary care settings), and test-retest reliability (Arroll et al., 2010). Additionally, the GAD-7 also has excellent internal consistency (Cronbach's α of 0.92) and test-retest reliability (Spitzer et al., 2006).

For participants in the earlier 8-week program, anxiety symptoms were measured at baseline and every 4 weeks through the program's end by the Generalized Anxiety Questionnaire-7 (GAD-7). For participants in the current 12-week program, anxiety symptoms were measured at baseline and every 2 weeks through the program's end by the GAD-7, a widely used instrument in outpatient and primary care settings to screen for the presence and severity of an anxiety disorder. The GAD-7 has excellent internal consistency and test-retest reliability (Löwe et al., 2008; Spitzer et al., 2006).

HRVB. Patients self-administered HRVB via the Meru Health app using a HeartMath Bluetooth photoplethysmography (PPG) sensor, which collects interbeat intervals (IBIs) from an earlobe. IBI data were then collected and sent to a secure database, which was then subsequently preprocessed and used for analyses.

HRVB Preprocessing. Preprocessing of the HRVB sessions was conducted using a standardized multi-step process provided by the RHRV package (Rodriguez-Linares et al., 2019). Before processing the data, we removed any HRVB session that was shorter than 180 seconds in duration in order to ensure consistency across calculations, and in order to ensure detection of low frequency signals. First, we extracted the non-interpolated instantaneous heart rate signal. Second, we used a filtering algorithm, which removed artifacts and ectopic beats based on whether they were 13 bpm away from the RR interval, the subsequent RR interval, or the average of 50 previous RR intervals, which was standard protocol within the preprocessing step in the RHRV package. Additionally, beats were removed if they were below 25 or above 200 bpm (Vila et al., Sep-Oct 1997). Third, we interpolated the heart rate signals that were removed during the filter process with a linear spline method using a frequency of 4 Hz (Hastie et al., 2013).

We chose to use PSD as the physiological marker for resonance zone breathing, as changes in PSD within the low frequency (LF) band have been shown to correspond to increasing baroreflex gains. Additionally, previous studies have used PSD to detect "Resonance Frequency" breathing within the LF range (Sakakibara et al., 2020; E. Vaschillo et al., 2002). Because we did not measure Resonance Frequency for each individual, but rather gave them a prescribed breathing pace (6 bpm), we will use the term "HRV Response" instead to indicate the PSD within a session's value. In order to ascertain the PSD for each session, we used the Lomb-Scargle periodogram. This allowed us to sum the PSD within a specific frequency band (i.e., within the resonance frequency band of 0.095 Hz – 0.112 Hz). Lastly, we normalized the sum of the PSD within the resonance frequency band by dividing the sum of this band with the sum of the PSD between 0.04Hz – 0.4Hz. This normalized sum of the PSD within a session was termed

“HRV Response,” and each recorded session had a single corresponding HRV Response score. See Figure 17 for a depiction of individuals with a high Normalized Peak, moderate Normalized Peak, and low Normalized Peak. Differences in HRV Response at the onset of treatment were estimated by using the PSD value of an individual’s very first HRVB session, whereas changes across treatment were estimated by extracting individual changes of PSD throughout the MHP. We did not include HRV Response values below 0.04 Hz for normalization because many HRVB sessions were not long enough to reliably calculate anything in the Very Low Frequency (VLF) range.

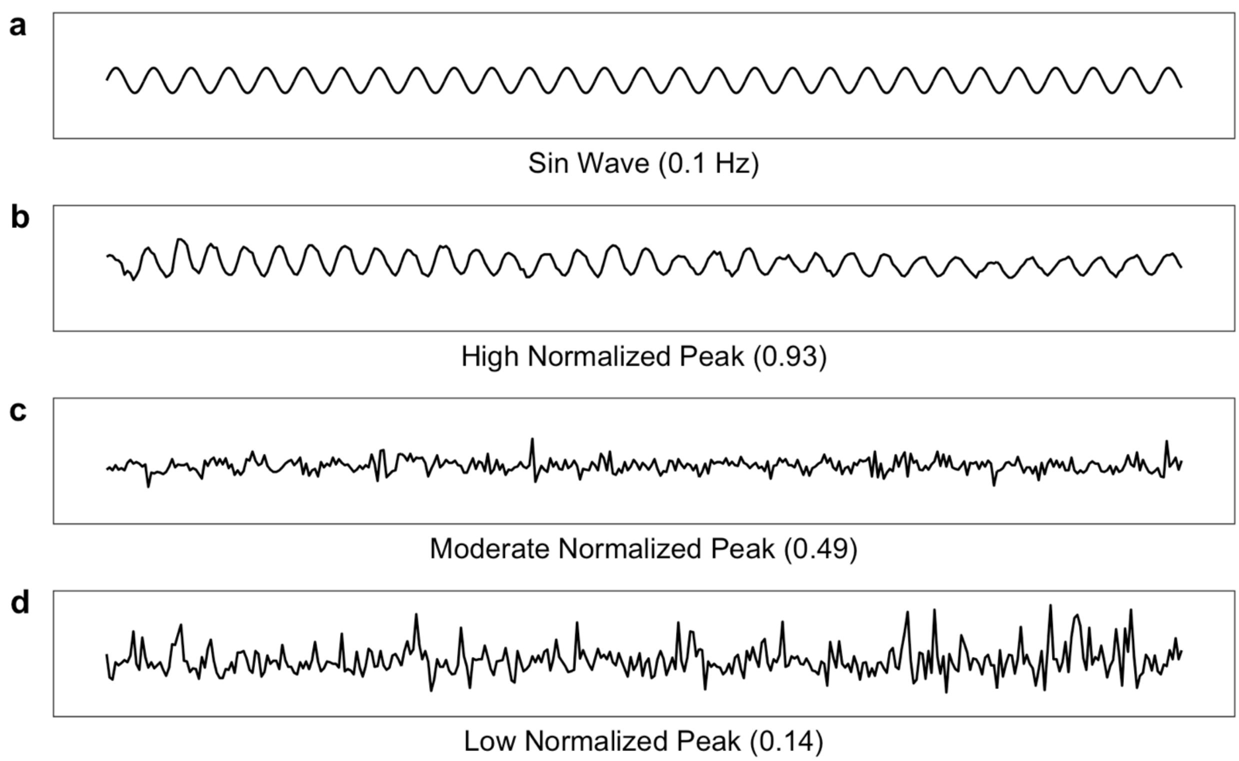

Figure 17
Figure 17a represents a sinusoidal wave with a frequency of 0.1Hz (6 breaths/minute). Figures 17b-1d represent the RR intervals of three separate HRVB sessions (same duration). Figure 17b is from an HRVB session with a high HRV Normalized Peak value at ~0.1 Hz; notice how similar it looks to the sinusoidal wave at 0.1 Hz. Figure 17c is from an HRVB session with a Normalized Peak value at 0.1 Hz; notice how it has some adherence to the 0.1 Hz frequency, but also much higher frequency noise. Figure 17d is from an HRVB session with a low Normalized Peak value at 0.1 Hz; notice how it has low adherence to the 0.1 Hz frequency, while having much higher frequency noise.

Covariates. Various patient demographics and clinical characteristics collected at baseline as well as engagement metrics collected during the MHP were examined as correlates of change in both depression and anxiety symptom response. The following covariates were included in all analyses (unless stated otherwise): age, gender, location of program (Finland or United States), length of program (8 weeks versus 12 weeks), baseline of PHQ-9 and GAD-7 scores, number of days active (e.g., engaging in any program practices, watching content, or messaging) with the MHP, and total time spent in HRVB sessions. We selected these covariates before our initial analyses, as opposed to using a stepwise regression for selecting these covariates, as using stepwise regression has led to overfitting, model instability, and issues replicating key findings (Babyak, 2004; Steyerberg et al., 2000). We chose these specific covariates for our analyses based on literature showing significant relationships with HRV as well as with depression and/or anxiety (Trivedi et al., 2006; Wong et al., 2001). See Table 26 for a list of descriptive statistics for each covariate.

Statistical Analysis

All statistical analyses were conducted in R (version 4.0.2). Statistical significance was defined using a critical alpha of 0.05. Descriptive statistics (i.e., n and percentages or mean and standard deviations) were calculated for each patient demographic and clinical variable, engagement characteristic, and each outcome variable (see Table 27). Outcome measures were analyzed using an intention-to-treat (ITT) analysis, in which all participants with outcome measures at baseline were included, regardless of intervention engagement or attrition, and last observation carried forward was used for the final PHQ-9 and GAD-7 scores (Streiner & Geddes, 2001).

For our analyses, we first ran a mixed-effects model using the lme4 package (Bates et al., 2014) in R to estimate individual trajectories in resonance zone breathing by nesting HRVB sessions within individuals. We modeled each individual's slope for normalized PSD within the resonance zone (0.095-0.112 Hz frequency band) over the course of the MHP by creating a "Days" variable, with 0 being the first day an individual used the HRVB device, as a predictor and PSD for each HRVB session as the predicted variable. Next, we centered the "Days" variable within each individual, which is suggested when using unbalanced time as a fixed- and random-effect in the model (Curran & Bauer, 2011). We then nested each of the sessions within an individual in a mixed-effects model, allowing us to extract an individual's slope (how their PSD in resonance zone breathing changed over the MHP; See Figure 18). Because these slopes were to be used in subsequent analyses, we did not adjust for covariates during this step, as we would be controlling for the same variables at a later stage in the analytic process. It should be noted that we removed individuals from analyses who had fewer than 6 HRVB sessions. This was done because the first few HRVB sessions were practice sessions in which the individual is orienting to the instructions, so the first few sessions might be sessions in which they are learning. In order to ascertain how physiological changes are related to clinical outcomes, multiple sessions over time are required, so we decided on a minimum of 6 HRVB sessions to examine this phenomenon.

We then used linear regression models to determine if changes in PHQ-9 and GAD-7 symptoms were related to these extracted PSD slopes, and PSD values from an individual's very first session.

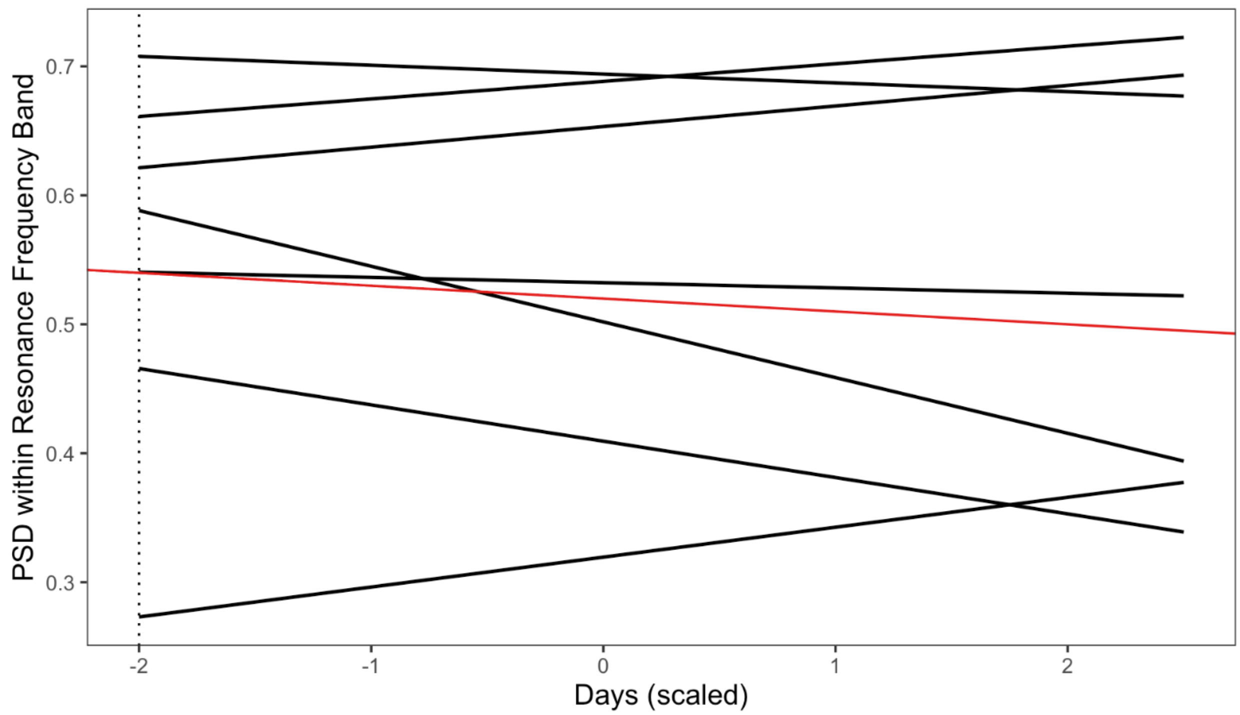

Figure 18.
An illustration that helps demonstrates how starting measurements and changes over time were determined for subsequent analyses. The starting measurements used in analyses were taken from the HRV Response values of each individual's baseline (i.e., first) HRVB session at the beginning of the intervention, and reflect initial starting ability. These points are where the black lines intersect with the vertical dotted line (Days = -2). The slopes represent changes in HRV Response throughout the intervention, and are estimates of changes (estimated using a mixed-effects model). The red line within the graph depicts the average slope and intercept of HRVB response throughout the MHP.

Results

Table 26.
Descriptive statistics for covariates.

Characteristic	Overall, N = 387
Total Active Days	38.03 (19.07)
Total HRVB Minutes	156.86 (135.89)
Total HRVB Sessions	22.21 (15.93)
Baseline PHQ-9	11.73 (6.14)
Baseline GAD-7	11.73 (4.64)

Completion Status	
Completer	361 / 387 (93%)
Dropout	26 / 387 (6.7%)
Age	39.66 (9.58)
Gender	
Female	307 / 387 (79%)
Male	80 / 387 (21%)
Program Location	
EU	70 / 387 (18%)
US	317 / 387 (82%)

Table 27
Descriptive statistics for variables of interest with variable mean (standard deviation).

Characteristic	Overall, N = 387
PHQ-9 Change Score	-5.02 (5.41)
GAD-7 Change Score	-4.95 (4.76)
HRVB PSD Baseline	0.55 (0.21)
HRVB PSD Slope	-0.01 (0.02)

MLM Descriptives.

The fixed-effects slope for scaled days on PSD of all individuals was $\hat{\beta} = -0.01$, 95% CI $[-0.02, -0.01]$, while the random-effects for the slope had a standard deviation of 0.16.

Changes Across HRV Response on PHQ-9 and GAD-7

First, we used the individual slopes of HRV Response as predictors of changes in depression (PHQ-9) and anxiety (GAD-7) symptom scores in separate models. Steeper positive slopes of HRV Response were significantly related to reductions in PHQ-9 scores across treatment ($b = -31.85$, 95% CI $[-50.83, -12.87]$, $t(376) = -3.30$, $p = .001$; see Table 28 & Figure 19). In order to determine whether the slope variable was predicting a significant portion of the variance, we removed it from the model to see if there was a significant change in R^2. When comparing a model with the slope included and with the slope excluded, there was a significant change ($\Delta R^2 = 0.02$, $p = .002$). Next, we wanted to determine whether changes in GAD-7 scores were predicted by changes in HRV Response slope. Steeper slopes of HRV Response were significantly related to reductions in GAD-7 scores across treatment ($b = -23.40$, 95% CI $[-39.91, -6.89]$, $t(376) = -2.79$, $p = .006$; see Table 29 & Figure 20). Again, we removed the slope value from the model to determine its explanatory value, and it did significantly add explanatory value to the model ($\Delta R^2 = 0.01$, $p = .013$).

Table 28.
Regression table showing the unstandardized and standardized betas for the HRVB practicing slope and intercepts extracted from the mixed-effects modeling on PHQ-9 scores.

	PHQ-9 Change Score (Baseline to Post-Intervention)				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
Intercept	7.35	0.66	3.95 – 10.75	0.27 – 1.05	**<0.001**
PSD Slope	-31.85	-0.13	-50.83 – -12.87	-0.22 – -0.05	**0.001**
PSD Baseline	-0.27	-0.01	-2.36 – 1.83	-0.09 – 0.07	0.803
Baseline PHQ9	-0.57	-0.65	-0.65 – -0.49	-0.74 – -0.55	**<0.001**
Baseline GAD7	0.05	0.04	-0.06 – 0.15	-0.05 – 0.13	0.376
Active Days	-0.08	-0.27	-0.11 – -0.05	-0.38 – -0.17	**<0.001**
Minutes Spent Practicing HRV	0.00	0.08	-0.00 – 0.01	-0.03 – 0.19	0.178

Age	-0.01	-0.02	-0.06 – 0.04	-0.10 – 0.06	0.636
Sex (Male)	0.63	0.12	-0.42 – 1.68	-0.08 – 0.31	0.238
Program Location (US)	-3.92	-0.72	-6.03 – -1.80	-1.11 – -0.33	**<0.001**
Program Length (8 week)	-2.85	-0.53	-5.02 – -0.68	-0.93 – -0.13	**0.010**
Observations	387				
R^2 / R^2 adjusted	0.429 / 0.414				

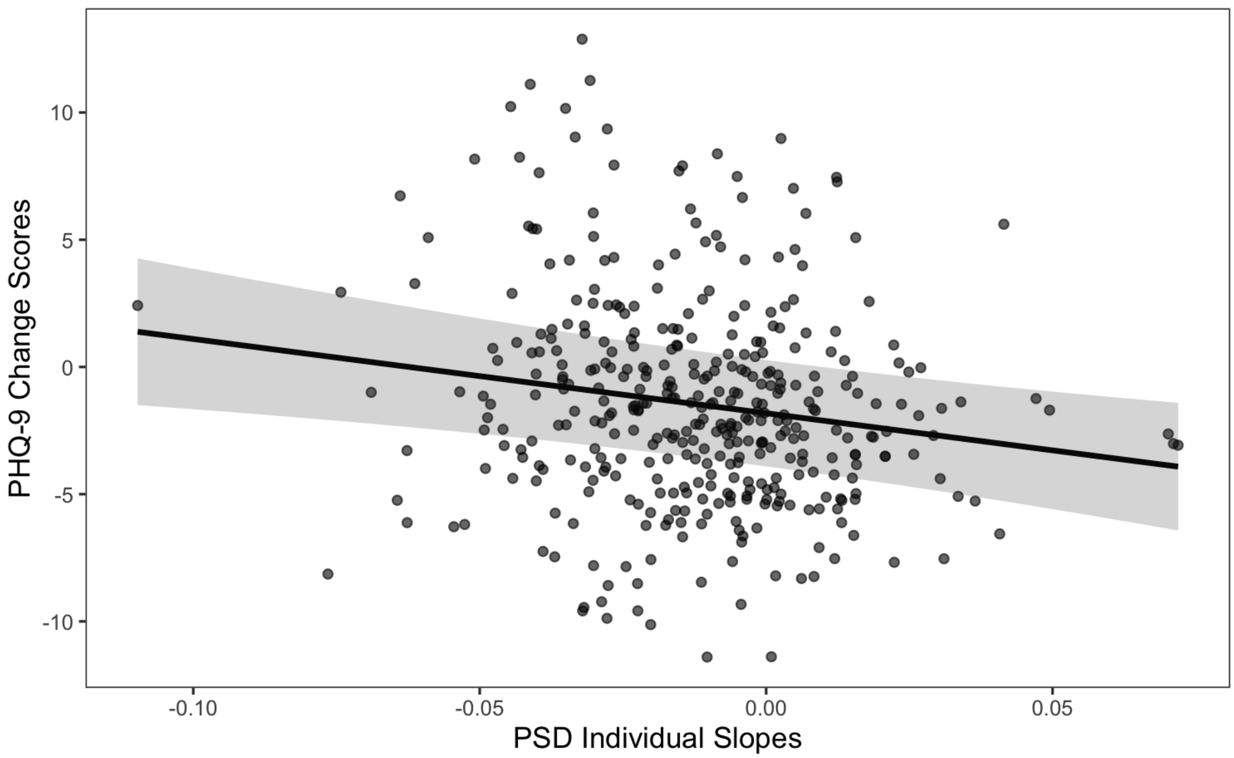

Figure 19.
A plot depicting the relationship between the extracted HRV Response Individual Slope values (see Figure 18) from the mixed-effects model and changes in PHQ-9 scores. Plotted points represent the partial residuals, in order to adjust for the effects of the other variables included in the model.

Table 29.
Regression table showing the unstandardized and standardized betas for the HRVB practicing slope and intercepts extracted from the mixed-effects modeling on GAD-7 scores.

	GAD7 Change Score (Baseline to Post-Intervention)				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
Intercept	8.26	0.61	5.31 – 11.22	0.23 – 0.99	**<0.001**
PSD Slope	-23.40	-0.11	-39.91 – -6.89	-0.19 – -0.03	**0.006**
PSD Baseline	-0.49	-0.02	-2.31 – 1.33	-0.10 – 0.06	0.598
Baseline PHQ9	0.07	0.09	-0.00 – 0.14	-0.00 – 0.18	0.061
Baseline GAD7	-0.64	-0.63	-0.73 – -0.55	-0.71 – -0.54	**<0.001**
Active Days	-0.07	-0.30	-0.10 – -0.05	-0.40 – -0.20	**<0.001**
Minutes Spent Practicing HRV	0.00	0.05	-0.00 – 0.01	-0.06 – 0.15	0.415
Age	-0.03	-0.05	-0.07 – 0.02	-0.13 – 0.03	0.219
Sex (Male)	0.15	0.03	-0.76 – 1.06	-0.16 – 0.22	0.741
Program Location (US)	-3.17	-0.67	-5.00 – -1.33	-1.05 – -0.28	**0.001**
Program Length (8 week)	-1.99	-0.42	-3.88 – -0.10	-0.81 – -0.02	**0.039**
Observations	387				
R^2 / R^2 adjusted	0.442 / 0.427				

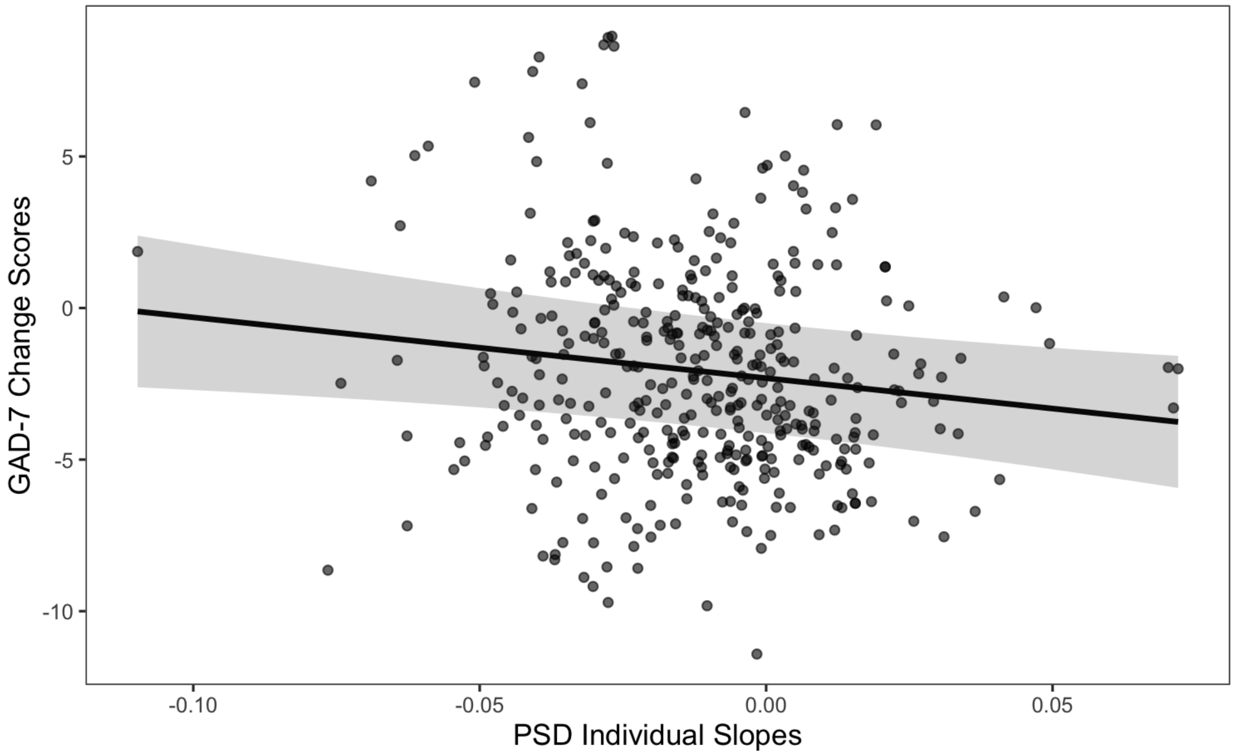

Figure 20
A plot depicting the relationship between the extracted HRV Response Individual Slope values (See Figure 18) from the mixed-effects model and changes in GAD-7 scores. Plotted points represent the partial residuals, in order to adjust for the effects of the other variables included in the model.

Differences in initial HRV Response on PHQ-9 and GAD-7

Individual baseline HRV Response values were not significantly associated with changes in PHQ-9 scores ($b = -0.27$, 95% CI $[-2.36, 1.83]$, $t(376) = -0.25$, $p = .803$; see Table 28), and they were also not significantly associated with changes in GAD-7 scores ($b = -0.49$, 95% CI $[-2.31, 1.33]$, $t(376) = -0.53$, $p = .598$; see Table 29).

Exploratory Analyses

We also wanted to determine whether different engagement factors would moderate the effect of slope on GAD-7 and PHQ-9 change scores. Specifically, we wanted to examine whether an individual's total number of HRVB sessions throughout the program and average HRVB session length throughout the program were related to clinical changes. We added the total number of HRVB sessions into the models as an interaction term with slope. For PHQ-9 change scores, neither the main effect of average session duration ($b = -0.01$, 95% CI [-0.05, 0.04], $t(374) = -0.27$, $p = .785$) nor the interaction term was significant ($b = -0.47$, 95% CI [-1.54, 0.59], $t(374) = -0.87$, $p = .383$). Similarly for GAD-7, the total number of HRVB sessions was not significant ($b = 0.04$, 95% CI [0.00, 0.07], $t(377) = 1.90$, $p = .058$), nor was the interaction term ($b = 0.31$, 95% CI [-0.61, 1.23], $t(377) = 0.66$, $p = .512$). We then switched out number of HRVB sessions in the models for average session duration to see if that (or its interaction term with slope) was related to PHQ-9 change scores, neither the main effect of average session duration ($b = -0.05$, 95% CI [-0.38, 0.27],$t(374) = -0.32$, $p = .749$) nor the interaction term was significant ($b = 5.42$, 95% CI [-4.17, 15.00], $t(374) = 1.11$, $p = .267$). Similarly for GAD-7, the average HRVB session duration was not significant ($b = -0.28$, 95% CI [-0.56, 0.00], $t(374) = -1.96$, $p = .051$), nor was the interaction term ($b = 2.23$, 95% CI [-6.06, 10.52], $t(377) = 0.53$, $p = .597$).

While we initially planned to perform analyses using the frequency band of 0.095-0.112 Hz, exploration of the data revealed that the vast majority of observations fell within the range of 0.095-0.112 Hz (See Figure 21). Therefore, we reran the analyses to determine how using this range would affect our overall results. Both the results remained the same; individual slopes were still significantly related to PHQ-9 scores ($b = -25.77$, 95% CI [-42.57, -8.98], $t(376) = -3.02$, $p = .003$) and GAD-7 scores ($b = -16.65$, 95% CI [-30.76, -2.53], $t(376) = -2.32$, $p = .021$). Baseline HRV Response values remained non-significantly associated with changes in PHQ-9 and GAD-7 scores.

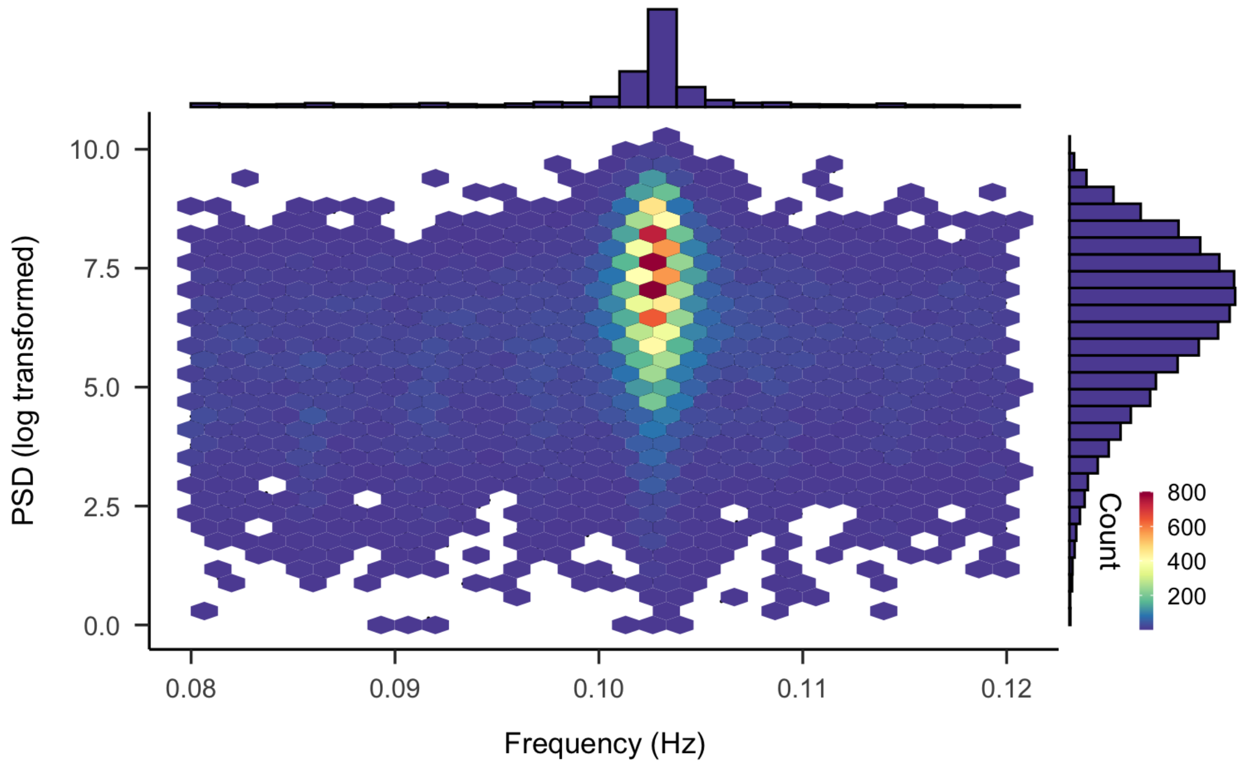

Figure 21.
A figure depicting the density of both the PSD and the max frequency for each session within the 0.08-0.12 Hz range. The vast majority of sessions have their peak frequency between 0.095-0.112, which is why we chose to restrict our exploratory analyses to this range.

Discussion

The current study investigated the association between PSD measured during 6 bpm HRVB practices and baseline to post-program changes in depressive and anxiety symptoms. More specifically, we examined whether slopes of HRV Response throughout treatment (i.e., changing PSD session values throughout treatment), and initial HRV Response values (measured by PSD) at the beginning of the intervention were associated with significant reductions in depressive and anxiety symptoms from baseline to post-treatment. The findings supported our hypothesis that changes in HRV Response would be related to reductions in mental health symptoms. When examining changes across treatment, we found that the more positive an individual's slope of their HRV Response across the MHP (i.e., improvements in adhering to the

6 bpm HRVB exercise), the greater the reductions in both PHQ-9 and GAD-7 scores from baseline to end of treatment. This effect remained significant even when using a more strict definition of HRV Response (0.095-0.112 Hz vs. 0.095-0.112 Hz). In contrast, initial HRV Response values assessed during the first HRVB session for each individual were not associated with reductions in PHQ-9 scores from baseline to the end of the program, nor changes in GAD-7 scores. These findings seem to indicate that repeated improvements in HRVB practices (as measured by PSD around the periodicity of 6 bpm) are related to clinical improvements, irrespective of initial starting ability. These findings are important, as these are the first to find that changes in a physiological marker have been shown to be associated with related reductions in depressive and anxiety symptoms.

These increases in HRV Response may be reflective of one or multiple mechanisms. Firstly, HRV Response may be reflective of how closely an individual is breathing to their resonance frequency. The closer one gets to breathing at resonance frequency, the higher the amplitude (Sakakibara et al., 2020). Stimulation at a frequency slightly off from resonance will still produce a high amplitude response but stimulation at an actual resonance will give a still greater amplitude (E. Vaschillo et al., 2002). Secondly, HRV Response can be impacted simply by how effectively an individual is adhering to breathing at 6 bpm throughout the entire session, since a cleaner “signal” would adhere to a better sinusoidal wave around 0.1 Hz. Thirdly, HRV Response could be also impacted by more long-term changes in baroreflex strength, caused by HRVB practices. Due to the amplitude being impacted by proximity to Resonance Breathing frequency, adherence to 6 bpm breathing, and long-term physiological factors (i.e., improvements in baroreflex reactivity), it is important to note that we are currently unable to differentiate between these based on the current study design.

These findings also confirm results from prior studies that have found significant associations between HRVB practice and improvements in mental health outcomes (Caldwell & Steffen, 2018), and provide important clinical implications for incorporating HRVB into mental health treatments. First, one's initial starting HRV Response was not associated with clinical gains, while *improvements* (irrespective of one's general ability) in HRV Response over time (i.e., slope of PSD) were associated with significant reductions in depression and anxiety symptoms across treatment. These findings suggest that not only is HRVB a malleable and trainable skill in mental health settings, but that physiological markers of change are associated with reductions in mental health symptoms. Second, these results indicate that improvement in patients' abilities to adhere to the 6 bpm HRVB exercises, regardless of their initial starting ability, are associated with improvements in mental health outcomes. Furthermore, because resonance zone breathing was significantly associated with decreases in depressive and anxiety symptoms above and beyond all covariates, including engagement and minutes spent practicing HRVB, it suggests that having participants breath at 6 bpm as a component of a mental health intervention likely has a generalizable impact on clinical outcomes regardless of total time spent engaging in program activities.

Because our sample contains some individuals with mild levels of both depression and anxiety (i.e., those with PHQ-9 and/or GAD-7 scores falling between 5-9), it is important to note the efficacy of HRVB training on populations that do not meet criteria for a diagnosable disorder. For example, a meta-analysis showed that there were no significant differences in effect size whether or not the underlying population had an anxiety diagnosis (Goessl et al., 2017). Similarly, another meta-analysis (Pizzoli et al., 2021) examining the effects of HRVB on depressive symptoms in non-clinical populations showed that there was a significant effect of

HRVB on depressive symptoms, regardless of symptom severity. Overall, these meta-analyses indicate that HRVB can be beneficial for individuals regardless of diagnostic status. However, it should be noted that, for our sample consists of of individuals who lie on a continuum of symptoms (from mild to severe), and would not be considered a “non-clinical” sample, so comparisons between our study and the ones mentioned above must be made with the knowledge that the findings from the meta-analyses may not have a direct comparison.

Limitations and Future Directions.

While this study contained several strengths including a robust sample of 387 individuals with a total of 16,109 HRVB sessions, our results should be interpreted within the context of several limitations. First, no clinical oversight was provided to help each study participant achieve resonance zone breathing, as the intervention was delivered remotely. Instead, each patient was taught to breathe at a rate of approximately 6 breaths per minute, which is not always the precise rate at which to achieve resonance zone breathing. While many individual’s natural resonance breathing is around 6 bpm, there is individual variability in ideal resonance frequency that could not be accounted for in the current study. Two studies have indicated that breathing at a personalized resonance frequency may produce better outcomes than breathing at resonance frequency plus one breath per minute or having all subjects breathe at 6 breaths per minute (Lin et al., 2014; Steffen et al., 2017). It is likely that personalizing the HRVB experience to an individual’s ideal resonance breathing rate would strengthen the relationship between these physiological changes and anxiety and depressive symptoms, but further studies must examine this.

Second, these are only findings from one sample of participants of a single mental health intervention; thus, replication of these findings and application to other types of samples are needed. In addition, our results only generalize to users of the MHP who practiced HRVB at least 6 times and should be interpreted as such; because patients who were excluded from analyses significantly differed in active days, completion rate, and age, this limits the generalizability of our study to those who are more likely to engage with and complete the program, and be younger. Third, the current study was unable to identify directionality of these effects, as it could be possible that individuals who improve at adhering to protocols, overall, are more likely to have improvements in mental health (which would, as a byproduct, lead to improved resonance frequency breathing), although we did adjust our analyses for total time of engagement with the MHP. Fourth, traditional HRVB is measured with both a PPG and a respiration monitor. Unfortunately, patients only had a PPG device to collect pulse information, and we were unable to use data from a respiration monitor to assess true resonance breathing. Lastly, this study did not include other treatment arms with interventions expected to have similar effects to HRVB on mental health outcomes, such as paced breathing at a different rate, which precludes the ability to conclude that beneficial effects came from resonance zone breathing specifically, rather than the alternate forms of adjunctive components within the MHP. For example, one study found that HRVB had no significant improvement when compared to a slowed breathing protocol in controlling physiological arousal and reducing anxiety (Wells et al., 2012), indicating that a simple breathing exercise may have comparable effectiveness without the need for HRVB devices, which would reduce the overall cost of treatment. Future studies should include an RCT component to determine the causal effects of improvements in HRVB training leads to reductions in mental health outcomes.

Conclusion

This study examined how changes in physiological markers of HRVB training at 6 bpm are related to changes in depression and anxiety symptoms. This is potentially the first step in identifying how a physiological metric could be useful for identifying those who are likely to experience reductions in anxiety and depressive symptoms. Changes in these physiological markers (as measured by PSD) were related to decreased anxiety and depressive symptoms. The current study is an important first step of determining if an indicator of physiological change is shown to track with overall changes in symptoms within a larger mHealth intervention.

Chapter 5

General Discussion

Aim 1 was to examine the sociodemographic factors associated with reductions in mental health symptoms in participants undergoing treatment in a DMHI. In order to do this, we conducted 2 separate studies. The first study examined the between-group differences in mental health symptoms across multiple timepoints for race & ethnicity, gender, and age participating in a DMHI. The second study consisted of examining differences in baseline scores for suicidality and trajectories over the intervention between gender expansive and cisgender participants. Results from study 1 revealed that Asian participants had significantly lower reported mental health symptoms compared with non-Asian participants. Additionally, women were likely to report higher depression and anxiety scores within the intervention compared to men and non-binary participants. Age was also significantly related to mental health symptoms across the span of the intervention, whereas older ages were related to significantly lower depression and anxiety scores than their younger counterparts.

In the second study, I found that gender expansive individuals had significantly higher baseline suicidality when compared to their cisgender counterparts. Despite this elevated baseline, they had a significantly steeper reduction in suicidality than cisgender participants. More interestingly, even accounting for other depressive symptoms, gender expansive individuals still showed reductions in suicidality, whereas for cisgender individuals this relationship was no longer observable.

Taken together, these results are rather promising. Despite Asian participants starting lower than their non-Asian counterparts, there was no interaction with time, indicating that everyone benefited equally from the intervention. This remained true for age and gender, meaning that no singular group benefited (or did worse) than the other groups. One group that did disproportionally benefit, however, was gender expansive participants, as it relates to reductions in suicidality. This is a very promising finding, as gender expansive individuals are highly likely to experience suicidality (Perez-Brumer et al., 2017). While we cannot go as far as to say this is an effective treatment for suicidal thoughts in this population, it does point to future research opportunities to determine if this could be the case. Additionally, to our knowledge, this is the first study that focused on sociodemographic factors associated with outcomes in a DMHI. Given the large sample size, this is truly a unique study.

Aim 2 examined an array of different engagement factors that are related to completion rates within a DMHI. The first portion examined the themes that arose in chat messages within a random subset of participants. These themes were related broad and encompassed discussing benefits of the intervention, issues with it, discussing everyday situations and feelings, and even asking for help with tech difficulties. This demonstrates that the chat function is being utilized for a multitude of reasons and is a useful aspect of the intervention. Further analyses revealed that messages from the therapist that are shorter are more likely to elicit a response than longer messages. Additionally, I employed a machine learning algorithm to determine which topics were predictive of dropping out of the intervention and found that therapist's requesting of phone calls and apologizing for difficulties were related to drop-out. Finally, using behavioral engagement measures and self-report measures, I determined that engagement measures are better predictive of completion than self-report measures. More specifically, the number of days spent active within the app was the strongest predictor of completion. This study was unique in that it utilized a mixed-methods approach to gather insights about text data and other engagement factors within the first week of the MHP. This study is unique, as, to my knowledge, no other study has examined the message data between therapist and client within the context of a company providing large-scale therapy to a broad range of individuals.

For study 3, our aim was to determine if there was an underlying biopsychological mechanism of change for improvements in depression and anxiety symptoms. We examined whether slopes of HRV Response throughout treatment, and initial HRV Response values at the beginning of the intervention were associated with significant reductions in depressive and anxiety symptoms from baseline to post-treatment. The findings supported our hypothesis that changes in HRV Response would be related to reductions in mental health symptoms. The important aspect of this research is that it demonstrates that within-person changes, independent of starting capabilities, are related to improvements in mental health. This research is novel for several reasons. First, this is one of the first studies that uses real-world data, not collected within a lab environment, but rather using HRVB as an adjunct to a therapy program. Second, this is the first study, to our knowledge, that has demonstrated how improvements to adherence in HRVB, as measured by a biomarker, is related to improvements in mental health symptoms.

There are several broader implications for this dissertation. The first broader implication is that it represents a significant shift in the way mental health services are being sought after and utilized by consumers. With the advent in mobile technology over the last decade, health applications and information are becoming far more ubiquitous than ever before. Individuals are often able to retrieve their personal health information from the comfort of their own home, and they are able to have greater control over their treatment. Mental health has been no different. There have been an abundance of applications and resources that have sprung up over the last decade aimed at ameliorating the effects of the burdens of mental illness, with varying degrees of efficacy. Recently, however, the overall quality and availability of these resources have dramatically increased. This is, at least in part, due to the recent COVID-19 pandemic. The global pandemic caused a shift in how many people navigate the world around them, as many people were unable to seek services in person. Additionally, the overall burden caused by the pandemic was related to an increase in mental health symptoms overall. Taken together, this means that there was increased need for mental health services and an overall increase in need for digital mental health services. This dissertation is reflective of this overall need, as the need for our understanding of how mental health tools translate into the digital world is sorely needed. Though there is research out there that has demonstrated that digital mental health tools can be as efficacious as in-person tools, nuanced research, such as understanding underlying mechanisms and impacts on different sociodemographic factors are sorely lacking. This dissertation hoped to add to this literature and broaden our understanding of how digital interventions work and who they might best work for.

Another implication that must be considered is the impact that DMHIs will have on accessibility. Individuals have a difficult time accessing in-person mental health services for a wide variety of reasons. Some people still face stigma within their family and cultures and are reluctant to seek out mental health services because they are worried about the impact it may have on their relationships (Benuto et al., 2019). Another reason individuals can have a difficult time accessing mental health services is due to physical barriers, such as living in rural areas or lack of transportation and accessible accommodations (Bornheimer et al., 2018; Summers-Gabr, 2020) (Bornheimer et al., 2018). Additionally, cost barriers (including transportation and taking time off work), have been shown to be a significant barrier for seeking in-person mental health services (Bornheimer et al., 2018; Summers-Gabr, 2020). DMHIs represent a unique opportunity to ameliorate the burden caused by these factors. Individuals are able to access a plethora of tools at any given moment (depending on the tool) and they can often do it within the comfort of their own home or wherever they find themselves needing services. This increase in accessibility not only helps everyone, but can actually be disproportionately beneficial for those who are marginalized and oppressed. For example, digital mental health care means that individuals seeking gender affirming care can find a therapist somewhere in their state without having to physically go to a location that might be too far out of their way. Additionally, individuals from cultures who are discouraged from seeking mental health treatment, such as Black, Latino, and Asian cultures, are better able to receive these services without people in their community knowing. This is important, because, often, people from marginalized communities have higher mental health burdens and could benefit most from the utilization of these services. The research in this dissertation highlights that digital mental health services can potentially be helpful for everyone, not just a select few groups of individuals.

Another implication of this research and digital mental health as a whole is its impact for mental health practitioners. Many of the tools available, such as HRVB devices, can be used as an adjunct to already existing in-person therapies that can be used at any point by their clients. Additionally, DMHIs like the Meru Health Program offer different models of therapy than have been traditionally utilized by therapists. For example, in traditional therapies, therapists often serve both as a supporting role and as someone who offers psychoeducation on different aspects of mental health and mental health interventions. In the MHP, much of the psychoeducation work is done by videos and exercises, leaving the therapists to mainly provide a supporting role in the client's journey through the intervention. While this might not be every therapist's preference, it does offer a new way for some therapists to operate who might prefer this model. Overall, DMHIs not only offer new ways of providing access to mental health services for clients, but offer potentially new and exciting avenues for mental health providers as well.

These studies have different limitations and strengths. One of the most notable limitations is the observational/quasi-experimental nature of these studies. Without a randomly selected control group, it is difficult to make any causal claims. Additionally, though this dissertation used real-world data, the generalizability of the data remains in question, as the selection process for the sample is inherently biased. Many of the participants that come to our program are referred either through a work program or through their health insurance. Many people do not have access to either of these options. Additionally, given that the intervention is only available in English, that limits the population to English speakers. Currently, people seeking treatment also have to be within the United States and have a smart phone that is connected to the internet. This means that our sample is not representative of a larger global population, nor representative of individuals that do not have access to smart phones and reliable internet. Finally, due to issues with confidentiality, I was the only one that was allowed to read through the chat messages between therapist and client; this meant that I was not able to get an interrater reliability index to check how accurate the coding of the chat messages was.

Despite all these limitations, there are several strengths to these studies. First, is the relatively large sample size. For the first study, we had a sample size of over 4,700 individuals. Given that we examined subsamples of individuals that are rarely represented in real-world research (and that have low penetrance in the population), this provided us with a large enough sample to make inferences about these populations that often are overlooked in research. Additionally, this is not an intervention that takes place in a controlled lab environment but is an actual intervention that thousands of individuals have participated in, meaning that our results are not hypothetical of what could be, but are reflective of the reality that people experience when undergoing the MHP.

Conclusion

For this dissertation, the goal was to examine the overall efficacy of a digital mental health intervention in the context of a real-world intervention using data collected by a mental health company. These findings helped us to better understand who this intervention was helpful to, how individuals engaged with different aspects of the intervention, and what a potential underlying mechanism of the intervention was. We found that, for depression and anxiety, the intervention was equally helpful across age, gender, and Race & Ethnicity. For suicidality, we found that gender expansive individuals receive greater reductions in suicidality than their cis gender counterparts. AIM 2. We also found that there was a potential underlying mechanism of change within the intervention. Increases in adherence to HRVB were related to decreases in mental health symptomatology. Overall, these findings are relatively novel, not only because DMHIs are so recent to the field, but also because many people have not examined how or why these interventions are effective.

One of the most novel aspects of this dissertation, however, is the nature in which the data was collected. Traditionally, data has been collected by researchers within academic or research institutions with small samples collected solely for the basis of research. This research is then used to inform mental health treatment in a broader way. However, it is rare to gain information about the efficacy of these interventions within a real-world context. The data collected for this dissertation was collected not by an academic or research institute, but rather collected by an entity that is conducting the intervention on a nation-wide level (within the U.S.)

and providing services to already over 10,000 individuals at the time of this writing. This is truly a relatively novel way of conducting research, because the data reflects the reality of how an intervention is being implemented, not just how it theoretically could be implemented. Additionally, this dissertation hopefully reflects the changing nature of our field, given that data is becoming much more readily available within real-world contexts. Hopefully, in the future, this type of data will be made readily available to researchers so that we can have a better understanding of how to help others benefit most from the research we conduct and the interventions we hope are impacting those suffering the most.

This research is important because it adds to a growing literature that mental health interventions are not limited to be what they always have been. The advent of new technology paves the way for more exciting ways that mental health work can expand who accesses services and what types of services they are accessing. The tools are no longer limited to a set of professionals that act can act as gatekeepers to this information. That is not to say that professionals will become obsolete, but rather this might actually *expand* the type of roles and opportunities that are available for practitioners. Overall, these advancements in technology benefit multiple groups of individuals. Clients can have more options and autonomy over their treatment, clinicians can occupy a wider array of roles, and researchers can have better access to data that has more generalizability and increases their understanding of mental health as a phenomenon. Ongoing empirical studies in conjunction with the tools and companies that are being formed will be necessary for guiding this new age of mental health care.

References

Alang, S. M. (2015). Sociodemographic disparities associated with perceived causes of unmet need for mental health care. *Psychiatric Rehabilitation Journal*, *38*(4), 293–299.

Alegría, M., Nakash, O., & NeMoyer, A. (2018). Increasing equity in access to mental health care: a critical first step in improving service quality. *World Psychiatry: Official Journal of the World Psychiatric Association* , *17*(1), 43–44.

American Association for Geriatric Psychiatry. (2004). Geriatrics and mental health—The facts. *American Association for Geriatric Psychiatry*.

Andrews, G., Basu, A., Cuijpers, P., Craske, M. G., McEvoy, P., English, C. L., & Newby, J. M. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. In *Journal of Anxiety Disorders* (Vol. 55, pp. 70–78). https://doi.org/10.1016/j.janxdis.2018.01.001

Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., Falloon, K., & Hatcher, S. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine*, *8*(4), 348–353.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*(3), 411–421.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. In *arXiv [stat.CO]*. arXiv. http://arxiv.org/abs/1406.5823

Beauchaine, T. P., Gatzke-Kopp, L., & Mead, H. K. (2007). Polyvagal Theory and developmental psychopathology: emotion dysregulation and conduct problems from preschool to adolescence. *Biological Psychology*, *74*(2), 174–184.

Beauchaine, T. P., & Thayer, J. F. (2015). Heart rate variability as a transdiagnostic biomarker of psychopathology. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *98*(2 Pt 2), 338–350.

Benuto, L. T., Gonzalez, F., Reinosa-Segovia, F., & Duckworth, M. (2019). Mental Health Literacy, Stigma, and Behavioral Health Service Use: the Case of Latinx and Non-Latinx Whites. *Journal of Racial and Ethnic Health Disparities*, *6*(6), 1122–1130.

Berger, S., Boettger, M. K., Tancer, M., Guinjoan, S. M., Yeragani, V. K., & Bär, K.-J. (2010). Reduced cardio-respiratory coupling indicates suppression of vagal activity in healthy relatives of patients with schizophrenia. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *34*(2), 406–411.

Borghouts, J., Eikey, E., Mark, G., De Leon, C., Schueller, S. M., Schneider, M., Stadnick, N., Zheng, K., Mukamel, D., & Sorkin, D. H. (2021). Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review. *Journal of Medical Internet Research*, *23*(3), e24387.

Bornheimer, L. A., Acri, M. C., Gopalan, G., & McKay, M. M. (2018). Barriers to Service Utilization and Child Mental Health Treatment Attendance Among Poverty-Affected Families. *Psychiatric Services* , *69*(10), 1101–1104.

Breslau, J., Kendler, K. S., Su, M., Gaxiola-Aguilar, S., & Kessler, R. C. (2005). Lifetime risk and persistence of psychiatric disorders across ethnic groups in the United States. *Psychological Medicine*, *35*(3), 317–327.

Buchhorn, R., Conzelmann, A., Willaschek, C., Störk, D., Taurines, R., & Renner, T. J. (2012). Heart rate variability and methylphenidate in children with ADHD. *Attention Deficit and Hyperactivity Disorders*, *4*(2), 85–91.

Caldwell, Y. T., & Steffen, P. R. (2018). Adding HRV biofeedback to psychotherapy increases heart rate variability and improves the treatment of major depressive disorder. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *131*, 96–101.

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A. J., & Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. *Depression and Anxiety*, *35*(6), 502–514.

Chalmers, J. A., Quintana, D. S., Abbott, M. J.-A., & Kemp, A. H. (2014). Anxiety Disorders are Associated with Reduced Heart Rate Variability: A Meta-Analysis. *Frontiers in Psychiatry / Frontiers Research Foundation*, *5*, 80.

Chekroud, A. M., Foster, D., Zheutlin, A. B., Gerhard, D. M., Roy, B., Koutsouleris, N., Chandra, A., Esposti, M. D., Subramanyan, G., Gueorguieva, R., Paulus, M., & Krystal, J. H. (2018). Predicting barriers to treatment for depression in a U.s. national sample: A cross-sectional, proof-of-concept study. *Psychiatric Services* , *69*(8), 927–934.

Comino, E. J., Silove, D., Manicavasagar, V., Harris, E., & Harris, M. F. (2001). Agreement in symptoms of anxiety and depression between patients and GPs: the influence of ethnicity. *Family Practice*, *18*(1), 71–77.

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica*, *144*(3), 288–299.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619.

D'Alfonso, S., Lederman, R., Bucci, S., & Berry, K. (2020). The Digital Therapeutic Alliance and Human-Computer Interaction. *JMIR Mental Health*, *7*(12), e21895.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269.

Dixon, L. B., Holoshitz, Y., & Nossel, I. (2016). Treatment engagement of individuals experiencing mental illness: review and update. *World Psychiatry: Official Journal of the World Psychiatric Association* , *15*(1), 13–20.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of Medical Internet Research*, *15*(11), e247.

Eckberg, D. L., & Sleight, P. (1992). *Human Baroreflexes in Health and Disease*. Clarendon Press.

Economides, M., Lehrer, P., Ranta, K., Nazander, A., Hilgert, O., Raevuori, A., Gevirtz, R., Khazan, I., & Forman-Hoffman, V. L. (2020). Feasibility and Efficacy of the Addition of Heart Rate Variability Biofeedback to a Remote Digital Health Intervention for Depression. *Applied Psychophysiology and Biofeedback*, *45*(2), 75–86.

Economides, M., Ranta, K., Nazander, A., Hilgert, O., Goldin, P. R., Raevuori, A., & Forman-Hoffman, V. (2019). Long-Term Outcomes of a Therapist-Supported, Smartphone-Based Intervention for Elevated Symptoms of Depression and Anxiety: Quasiexperimental, Pre-Postintervention Study. In *JMIR mHealth and uHealth* (Vol. 7, Issue 8, p. e14284). https://doi.org/10.2196/14284

Edman, J. L., Danko, G. P., Andrade, N., McArdle, J. J., Foster, J., & Glipa, J. (1999). Factor structure of the CES-D (Center for Epidemiologic Studies Depression Scale) among

Filipino-American adolescents. *Social Psychiatry and Psychiatric Epidemiology*, *34*(4), 211–215.

Fernández-Alvarez, J., Grassi, M., Colombo, D., Botella, C., Cipresso, P., Perna, G., & Riva, G. (2022). Efficacy of bio- and neurofeedback for depression: a meta-analysis. *Psychological Medicine*, *52*(2), 201–216.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry: Official Journal of the World Psychiatric Association* , *16*(3), 287–298.

Fumarco, L., Harrell, B., Button, P., Schwegman, D., & Dils, E. (2020). *Gender Identity, Race, and Ethnicity-based Discrimination in Access to Mental Health Care: Evidence from an Audit Correspondence Field Experiment* (No. 28164). National Bureau of Economic Research. https://doi.org/10.3386/w28164

Goessl, V. C., Curtiss, J. E., & Hofmann, S. G. (2017). The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. *Psychological Medicine*, *47*(15), 2578–2586.

Goldin, P. R., Lindholm, R., Ranta, K., Hilgert, O., Helteenvuori, T., & Raevuori, A. (2019). Feasibility of a Therapist-Supported, Mobile Phone–Delivered Online Intervention for Depression: Longitudinal Observational Study. *JMIR Formative Research*, *3*(1), e11509.

Goodwin, R. D., Weinberger, A. H., Kim, J. H., Wu, M., & Galea, S. (2020). Trends in anxiety among adults in the United States, 2008-2018: Rapid increases among young adults. *Journal of Psychiatric Research*, *130*, 441–446.

Hasin, D. S., Sarvet, A. L., Meyers, J. L., Saha, T. D., Ruan, W. J., Stohl, M., & Grant, B. F. (2018). Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. *JAMA Psychiatry* , *75*(4), 336–346.

Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

Heck, N. C., Mirabito, L. A., LeMaire, K., Livingston, N. A., & Flentje, A. (2017). Omitted data in randomized controlled trials for anxiety and depression: A systematic review of the inclusion of sexual orientation and gender identity. *Journal of Consulting and Clinical Psychology*, *85*(1), 72–76.

Henry, B. L., Minassian, A., Paulus, M. P., Geyer, M. A., & Perry, W. (2010). Heart rate variability in bipolar mania and schizophrenia. *Journal of Psychiatric Research*, *44*(3), 168–176.

Hyde, J. S., Mezulis, A. H., & Abramson, L. Y. (2008). The ABCs of depression: integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. *Psychological Review*, *115*(2), 291–313.

Kaltenthaler, E., Sutcliffe, P., Parry, G., Beverley, C., Rees, A., & Ferriter, M. (2008). The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. *Psychological Medicine*, *38*(11), 1521–1530.

Karyotaki, E., Ebert, D. D., Donkin, L., Riper, H., Twisk, J., Burger, S., Rozental, A., Lange, A., Williams, A. D., Zarski, A. C., Geraedts, A., van Straten, A., Kleiboer, A., Meyer, B., Ünlü Ince, B. B., Buntrock, C., Lehr, D., Snoek, F. J., Andrews, G., … Cuijpers, P. (2018). Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. *Clinical Psychology Review*, *63*, 80–92.

Kemp, A. H., Koenig, J., & Thayer, J. F. (2017). From psychological moments to mortality: A multidisciplinary synthesis on heart rate variability spanning the continuum of time. *Neuroscience and Biobehavioral Reviews*, *83*, 547–567.

Kemp, A. H., & Quintana, D. S. (2013). The relationship between mental and physical health: insights from the study of heart rate variability. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *89*(3), 288–296.

Kemp, A. H., Quintana, D. S., Gray, M. A., Felmingham, K. L., Brown, K., & Gatt, J. M. (2010). Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. *Biological Psychiatry*, *67*(11), 1067–1074.

Kessler, R. C., Berglund, P. A., Bruce, M. L., Koch, J. R., Laska, E. M., Leaf, P. J., Manderscheid, R. W., Rosenheck, R. A., Walters, E. E., & Wang, P. S. (2001). The prevalence and correlates of untreated serious mental illness. *Health Services Research*, *36*(6 Pt 1), 987–1007.

Kirby, J. B., Zuvekas, S. H., Borsky, A. E., & Ngo-Metzger, Q. (2019). Rural Residents With Mental Health Needs Have Fewer Care Visits Than Urban Counterparts. *Health Affairs* , *38*(12), 2057–2060.

Kirvin-Quamme, A., Kissinger, J., Quinlan, L., Montgomery, R., Chernenok, M., Pirner, M. C., Pajarito, S., Rapoport, S., Wicks, P., Darcy, A., Greene, C. J., & Robinson, A. (2023). Common practices for sociodemographic data reporting in digital mental health intervention research: A scoping review. In *bioRxiv*. https://doi.org/10.1101/2023.07.24.23293020

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*(9), 606–613.

Lee, J.-S., Kim, B., Hong, Y., & Joo, Y. H. (2012). Heart rate variability in the subsyndromal depressive phase of bipolar disorder. *Psychiatry and Clinical Neurosciences*, *66*(4), 361–366.

Lehrer, P. (2013). How Does Heart Rate Variability Biofeedback Work? Resonance, the Baroreflex, and Other Mechanisms. *Biofeedback and Self-Regulation*, *41*(1), 26–31.

Lehrer, P., & Gevirtz, R. (2014). Heart rate variability biofeedback: how and why does it work? *Frontiers in Psychology*, *5*, 756.

Lehrer, P., Kaur, K., Sharma, A., Shah, K., Huseby, R., Bhavsar, J., Sgobba, P., & Zhang, Y. (2020). Heart Rate Variability Biofeedback Improves Emotional and Physical Health and Performance: A Systematic Review and Meta Analysis. *Applied Psychophysiology and Biofeedback*, *45*(3), 109–129.

Lehrer, P. M., Vaschillo, E., Vaschillo, B., Lu, S.-E., Eckberg, D. L., Edelberg, R., Shih, W. J., Lin, Y., Kuusela, T. A., Tahvanainen, K. U. O., & Hamer, R. M. (2003). Heart rate variability biofeedback increases baroreflex gain and peak expiratory flow. *Psychosomatic Medicine*, *65*(5), 796–805.

Lewandowski, G. W., Jr, & Strohmetz, D. B. (2009). Actions can speak as loud as words: Measuring behavior in psychological science. *Social and Personality Psychology Compass*, *3*(6), 992–1002.

Lin, I. M., Tai, L. Y., & Fan, S. Y. (2014). Breathing at a rate of 5.5 breaths per minute with equal inhalation-to-exhalation ratio increases heart rate variability. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *91*(3), 206–211.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, *46*(3), 266–274.

Mather, M., & Thayer, J. (2018). How heart rate variability affects emotion regulation brain networks. *Current Opinion in Behavioral Sciences*, *19*, 98–104.

Mimno, D., & Lee, M. (2014). *Low-dimensional Embeddings for Interpretable Anchor-based Topic Inference*. 1319–1328.

Nashiro, K., Yoo, H. J., Min, J., Cho, C., Nasseri, P., Zhang, Y., Lehrer, P., Thayer, J. F., & Mather, M. (2022). Effects of a randomised trial of 5-week heart rate variability biofeedback intervention on mind wandering and associated brain function. *Cognitive, Affective & Behavioral Neuroscience*, *22*(6), 1349–1357.

Novella, J. K., Ng, K.-M., & Samuolis, J. (2022). A comparison of online and in-person counseling outcomes using solution-focused brief therapy for college students with anxiety. *Journal of American College Health: J of ACH*, *70*(4), 1161–1168.

Novick, D., Montgomery, W., Aguado, J., Kadziola, Z., Peng, X., Brugnoli, R., & Haro, J. M. (2013). Which somatic symptoms are associated with an unfavorable course in Asian patients with major depressive disorder? *Journal of Affective Disorders*, *149*(1–3), 182–188.

Obikane, E., Sasaki, N., Imamura, K., Nozawa, K., Vedanthan, R., Cuijpers, P., Shimazu, T., Kamada, M., Kawakami, N., & Nishi, D. (2022). Usefulness of Implementation Outcome Scales for Digital Mental Health (iOSDMH): Experiences from Six Randomized Controlled Trials. *International Journal of Environmental Research and Public Health*, *19*(23). https://doi.org/10.3390/ijerph192315792

Olfson, M., Blanco, C., Wall, M. M., Liu, S.-M., & Grant, B. F. (2019). Treatment of Common Mental Disorders in the United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions-III. *The Journal of Clinical Psychiatry*, *80*(3). https://doi.org/10.4088/JCP.18m12532

Parker, G., & Brotchie, H. (2010). Gender differences in depression. *International Review of Psychiatry* , *22*(5), 429–436.

Perez-Brumer, A., Day, J. K., Russell, S. T., & Hatzenbuehler, M. L. (2017). Prevalence and Correlates of Suicidal Ideation Among Transgender Youth in California: Findings From a Representative, Population-Based Sample of High School Students. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(9), 739–746.

Pizzoli, S. F. M., Marzorati, C., Gatti, D., Monzani, D., Mazzocco, K., & Pravettoni, G. (2021). A meta-analysis on heart rate variability biofeedback and depressive symptoms. *Scientific Reports*, *11*(1), 6650.

Progovac, A. M., Mullin, B. O., Dunham, E., Reisner, S. L., McDowell, A., Sanchez Roman, M. J., Dunn, M., Telingator, C. J., Lu, F. Q., Breslow, A. S., Forstein, M., & Cook, B. L. (2020). Disparities in Suicidality by Gender Identity Among Medicare Beneficiaries. *American Journal of Preventive Medicine*, *58*(6), 789–798.

Quintana, D. S., McGregor, I. S., Guastella, A. J., Malhi, G. S., & Kemp, A. H. (2013). A meta-analysis on the impact of alcohol dependence on short-term resting-state heart rate variability: implications for cardiovascular risk. *Alcoholism, Clinical and Experimental Research*, *37 Suppl 1*(s1), E23-9.

R Foundation for Statistical Computing, R. (2018). R: A language and environment for statistical computing. *RA Lang Environ Stat Comput*.

Reisner, S. L., Katz-Wise, S. L., Gordon, A. R., Corliss, H. L., & Austin, S. B. (2016). Social Epidemiology of Depression and Anxiety by Gender Identity. *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine*, *59*(2), 203–208.

Reisner, S. L., Vetters, R., Leclerc, M., Zaslow, S., Wolfrum, S., Shumer, D., & Mimiaga, M. J. (2015). Mental health of transgender youth in care at an adolescent urban community health center: a matched retrospective cohort study. *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine*, *56*(3), 274–279.

Reynolds, W. M., & Kobak, K. A. (1995). Reliability and validity of the Hamilton Depression Inventory: A paper-and-pencil version of the Hamilton Depression Rating Scale Clinical Interview. *Psychological Assessment*, *7*(4), 472–483.

Rodriguez-Linares, L., Vila, X., Lado, M. J., Mendez, A., Otero, A., Garcia, C. A., & Others. (2019). RHRV: heart rate variability analysis of ECG data. *R Package Version*, *4*(5).

Sakakibara, M., Kaneda, M., & Oikawa, L. O. (2020). Efficacy of Paced Breathing at the Low-frequency Peak on Heart Rate Variability and Baroreflex Sensitivity. *Applied Psychophysiology and Biofeedback*, *45*(1), 31–37.

Schueller, S. M., Hunter, J. F., Figueroa, C., & Aguilera, A. (2019). Use of Digital Mental Health for Marginalized and Underserved Populations. *Current Treatment Options in Psychiatry*, *6*(3), 243–255.

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*, 258.

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology*, *5*, 1040.

Shinba, T., Kariya, N., Matsui, Y., Ozawa, N., Matsuda, Y., & Yamamoto, K.-I. (2008). Decrease in heart rate variability response to task is related to anxiety and depressiveness in normal subjects. *Psychiatry and Clinical Neurosciences*, *62*(5), 603–609.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

Steffen, P. R., Austin, T., DeBarros, A., & Brown, T. (2017). The Impact of Resonance Frequency Breathing on Measures of Heart Rate Variability, Blood Pressure, and Mood. *Frontiers in Public Health*, *5*, 222.

Steyerberg, E. W., Eijkemans, M. J., Van Houwelingen, J. C., Lee, K. L., & Habbema, J. D. (2000). Prognostic models based on literature and individual patient data in logistic regression analysis. *Statistics in Medicine*, *19*(2), 141–160.

Streiner, D., & Geddes, J. (2001). Intention to treat analysis in clinical trials when there are missing data. *Evidence-Based Mental Health*, *4*(3), 70–71.

Summers-Gabr, N. M. (2020). Rural–urban mental health disparities in the United States during COVID-19. *Psychological Trauma: Theory, Research, Practice and Policy*, *12*(1), S222–S224.

Sutarto, A. P., Wahab, M. N. A., & Zin, N. M. (2012). Resonant breathing biofeedback training for stress reduction among manufacturing operators. *International Journal of Occupational Safety and Ergonomics: JOSE*, *18*(4), 549–561.

Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, *37*(2), 141–153.

Tiwari, S. K., & Wang, J. (2006). The epidemiology of mental and substance use-related disorders among white, Chinese, and other Asian populations in Canada. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *51*(14), 904–912.

Torous, J., & Haim, A. (2018). Dichotomies in the Development and Implementation of Digital Mental Health Tools. *Psychiatric Services* , *69*(12), 1204–1206.

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., Norquist, G., Howland, R. H., Lebowitz, B., McGrath, P. J., Shores-Wilson, K., Biggs, M. M., Balasubramani, G. K., & Fava, M. (2006). Evaluation of Outcomes With Citalopram for Depression Using Measurement-Based Care in STAR*D: Implications for Clinical Practice. *American Journal of Psychiatry*, *163*(1), 28–40.

van Dis, E. A. M., van Veen, S. C., Hagenaars, M. A., Batelaan, N. M., Bockting, C. L. H., van den Heuvel, R. M., Cuijpers, P., & Engelhard, I. M. (2020). Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders: A systematic review and meta-analysis. *JAMA Psychiatry* , *77*(3), 265–273.

Vaschillo, E. G., Vaschillo, B., & Lehrer, P. M. (2006). Characteristics of resonance in heart rate variability stimulated by biofeedback. *Applied Psychophysiology and Biofeedback*, *31*(2), 129–142.

Vaschillo, E., Lehrer, P., Rishe, N., & Konstantinov, M. (2002). Heart rate variability biofeedback as a method for assessing baroreflex function: a preliminary study of resonance in the cardiovascular system. *Applied Psychophysiology and Biofeedback*, *27*(1), 1–27.

VERBI Software. (2022). *MAXQDA*. Berlin, Germany: VERBI Software. maxqda.com

Vila, J., Palacios, F., Presedo, J., Fernandez-Delgado, M., Felix, P., & Barro, S. (Sep-Oct 1997). Time-frequency analysis of heart-rate variability. *IEEE Engineering in Medicine and Biology Magazine: The Quarterly Magazine of the Engineering in Medicine & Biology Society*, *16*(5), 119–126.

von Brachel, R., Teismann, T., Feider, L., & Margraf, J. (2019). Suicide ideation as a predictor of treatment outcomes in cognitive-behavioral therapy for unipolar mood disorders. *International Journal of Clinical and Health Psychology: IJCHP*, *19*(1), 80–84.

Walker, E. R., Cummings, J. R., Hockenberry, J. M., & Druss, B. G. (2015). Insurance status, use of mental health services, and unmet need for mental health care in the United States. *Psychiatric Services* , *66*(6), 578–584.

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 629–640.

Wells, R., Outhred, T., Heathers, J. A. J., Quintana, D. S., & Kemp, A. H. (2012). Matter over mind: a randomised-controlled trial of single-session biofeedback training on performance anxiety and heart rate variability in musicians. *PloS One*, *7*(10), e46597.

Wong, N. D., Pio, J., Valencia, R., & Thakal, G. (2001). Distribution of C-reactive protein and its relation to risk factors and coronary heart disease risk estimation in the National Health and Nutrition Examination Survey (NHANES) III. *Preventive Cardiology*, *4*(3), 109–114.

Yoo, H. J., Nashiro, K., Min, J., Cho, C., Bachman, S. L., Nasseri, P., Porat, S., Dutt, S., Grigoryan, V., Choi, P., Thayer, J. F., Lehrer, P. M., Chang, C., & Mather, M. (2022). Heart rate variability (HRV) changes and cortical volume changes in a randomized trial of five weeks of daily HRV biofeedback in younger and older adults. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *181*, 50–63.

Zhang, J., & Norvilitis, J. M. (2002). Measuring Chinese psychological well-being with Western developed instruments. *Journal of Personality Assessment*, *79*(3), 492–511.

Zhang, Z., Zhang, L., Zhang, G., Jin, J., & Zheng, Z. (2018). The effect of CBT and its modifications for relapse prevention in major depressive disorder: a systematic review and meta-analysis. *BMC Psychiatry*, *18*(1), 50.