Assessing the utility of digital health technology to improve our capacity to assess and intervene in depression

Abhishek Pratap

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Sean Mooney, Chair

Patrick J. Heagerty

Patricia Areán

Program Authorized to Offer Degree:

Biomedical Informatics and Medical Education

University of Washington

Abstract

Assessing the utility of digital health technology to improve our capacity to assess and intervene in depression

Abhishek Pratap

Chair of the Supervisory Committee:
Dr. Sean Mooney
Professor, Biomedical Informatics and Medical Education
Chief Research Information Officer (CRIO), UW Medicine
Biomedical Informatics and Medical Education

When it comes to mental health, no country is considered developed. In the last decade, the burden of mental health disorders (MHD) has risen in all countries due to disparities in timely diagnosis and access to evidence-based treatments. Additionally, scientists, are still conducting research to understand the underlying mechanisms behind MHD. Part of the problem is that measures of symptom severity are all based on self-reports by patients and clinician observation often resulting in an imprecise measurement of MHD. Those that are more objective (e.g. MRI) are costly and not widely available, nor are they ecologically valid measures of behavior. Additionally, in-clinic assessments tend to be episodic and often miss capturing the lived experience of disease over time

including the potential impact of social and environmental factors that are suspected to be linked to neurodevelopmental and psychological processes. To improve long term outcomes in MHD, there is a critical need to develop new ways to objectively assess specific underlying constructs of behavior patterns linked with neuropsychiatric conditions. The pervasive network of smartphones offers researchers a unique opportunity to study MH at a population scale and at a fraction of the cost of traditional clinical research. The high-frequency daily usage of smartphones also provides new ways to capture the individualized momentary experience of living with mental health issues based on "real-world data" (RWD) in an objective, momentary and nonreactive way.

The principal findings of this dissertation research show the feasibility of utilizing smartphones to reach, enroll and engage a diverse and nationally representative population as well as the potential of using RWD in predicting mental health outcomes. The RWD collected from more than 2000 participants showed notable inter-/intra-person heterogeneity highlighting the challenges of developing a robust cohort level machine learning model to predict depression. However, personalized N-of-1 models show the promise of "precision digital psychiatry" by assessing an individual's drifts from their own average "digital behavior" as a more reliable predictor of a person's daily mood. Of note, participant enrollment and retention in large-scale digital health research studies remains a significant challenge. Cross study analysis using data from >100,000 participants showed significant underlying biases in technology access and utilization based on participants' demographics that could impact the generalizability of the statistical inference drawn. In addition, the results from a survey-based study on a large and diverse sample show growing concerns among the general public about the security and privacy of their digital data which if left

unaddressed can negatively influence people's decision to participate and share data in digital health research.

These findings are contemporary and extend the on-going efforts to objectively evaluate the potential fit of technology in psychiatry in engaging the general population to monitor their mental health in the real world outside the clinic. However, while the technology shows the promise to move the psychiatric research from subjective to objective measures, episodic to continuous monitoring, provider-based to ubiquitous and reactive to proactive care; accomplishing these goals does come with measurable challenges. Further research is needed to develop robust and validated digital biomarkers of behavioral health. This includes large scale behavioral phenotyping studies ($N > 100,000$) that are powered to detect the association between RWD and behavioral anomalies, the ability to integrate RWD across similar studies, improve equitable utilization of technology across a diverse and representative population and address people's concerns about data security and privacy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

"Some memories are unforgettable, remaining ever vivid and heartwarming!"

Joseph B. Wirthlin

My journey exploring the use of smartphone technology for remote assessment of health especially in mental health has been very special. The invaluable experience I gained in the last five years working with so many amazing colleagues, collaborators and mentors has transformed my career and will continue to reshape it in the future too. I consider myself extremely fortunate to have the opportunities to work on many interdisciplinary projects that I could not have thought about five years ago. And for that, I am truly indebted to so many who over the last five years believed in me.

I would like to thank my committee members Dr. Pat Arean, Dr. Patrick Heagerty, Dr. Sean Mooney and Dr. James Fogarty for continued guidance and mentorship. My sincere gratitude to Dr. Arean for the countless number of meetings and discussions over the last four years. You helped me realize the potential of digital technology in bringing a sea change in mental health research. Thank you, Dr. Heagerty, for helping me keep the statistical analysis rigorous. And thank you Dr. Stephen Friend and Dr. Andrew Trister for helping me distill and focus my meandering thoughts about pursuing a Ph.D. five years ago. Without your mentorship and support, my journey would not have been the same. Also my immense gratitude to Sage Bionetworks for the tremendous support during my research. I would be remiss not to mention how lucky I have been to have amazing collaborators in Dr. John Torous, Dr. Honor Hsin, Dr. David Mohr, Theresa Nguyen alongside a fantastic group of admins Jaden Duffy, Sarah Morrow, Jill Fulmore, David Lahiti and Diane Gary who made the paperwork and meeting scheduling a

breeze. It was a pleasure to work with you all and the experience of a lifetime that I will cherish for a long time. Thank you.

Pursuing my work and research together will not be possible if not for my friends and family. Thank you, Mumma Papa and Richa for everything. While I didn't always say what I should have and maybe words are not enough to express my feelings, I will forever be grateful for all the things you did. Richa thank you for always being there for me, making sure I completed my assignments as much as I dreaded and for being a super mom to Advik. Finally, one common string that binds all who have touched and positively impacted my experience is being generous and to pay it forward. Thank you once again for being there for me and I will do my best to pay it forward.

DEDICATION

I dedicate my dissertation work to my late grandfather Lala Asthami Chand Ji, to my parents Manju and Ganga Pratap Agrawal and to thousands of individuals who participated and contributed data in remote online research.

Chapter 1. INTRODUCTION

"Let no one deceive himself: trying to understand another human being's emotional life is fraught with potential error … As intuition is greatly influenced by one's own prejudices and needs, it lends an air of deceptive yet powerful plausibility. This is especially worrying as we have no objective yardstick for this confidence."

– Emil Kraepelin, The Manifestations of Insanity, 1920[1]

1.1 MENTAL HEALTH

Our psychological, social and emotional wellbeing i.e "Mental health" affects how we think, feel and behave in our daily lives from childhood to adulthood. Any prolonged adverse impact on mental functioning can directly impact our way of thinking, how we relate to others and potentially impede day-to-day functioning. Mental health is therefore central to human health and should be considered and treated at par with other diseases. However, most people with mental health disorders (MHD) are not able to receive minimally adequate and timely care in both high- and low-resource settings alike[2]. It is estimated 29% of the people will experience a mental health-related disorder (MHD) in their lifetime[3]. WHO estimates show Mental Health disorders (MHD) are amongst the most burdensome diseases worldwide affecting the quality of life and with billions of dollars in lost earnings per year. In the US, the cost of treating MHD is among the top 5 most expensive health conditions[4]. In 2013, $201 billion were spent providing care to people with MHD[5] and this figure is expected to reach $237 billion by 2020[6] with an expected annual increase of 2.8%; higher than Oncology at 1.3% and the national average[7] of 1.8%. MHD are also the leading cause of disease burden[8,9] with the 12-month prevalence estimated[10–13] to be around 18-25%. This means close to 1 in 5 Americans (44.7 million in 2016[14]) will suffer from some form of mental illness or its sequelae every year. Given the broad impact of MHD on public health and along with the economic costs; improving long-term outcomes for MHD by accurate early diagnosis, and interventions has been a focus of national and international agencies. World Health Organization(WHO) in it's Mental Health Action Plan 2013–2020[15] has also recommended the use of mobile-based technology to improve MHD outcomes "the promotion of self-care, for instance, through the use of electronic and mobile health technologies". Mental health is now a

key part of the WHO Sustainable Development Goals[16] (SDGs) for the year 2030. The recent Lancet commission's 2018 report[17] highlights the key four key pillars for mental health with the potential to bring a sea change in improving outcomes in MHD including the use of technology-aided solutions.

a.) *Mental health issues are impacting all countries alike regardless of their socioeconomic status:* Estimates show the prevalence of depression[18] and death by suicides[19] are in parity across the developed and developing world. There exists a significant gap in the need and availability of care for MHD. Up to 55% of people in developed countries and a staggering 85% in the developing countries are not able to get the needed treatment[20] despite the fact that there are many evidence-based treatments for most MHDs[21]. Furthermore, several disparities exist in access, service utilization and quality of care received by ethnic minorities and people living in remote areas in both low[22] and high[23,24] resource settings. Developing countries like India with the world's second-largest population has an estimated 4000 psychiatrists which translates to just 3 psychiatrists for every 1 million people with 75% of this workforce work in urban areas, thus addressing only 31% of the country's population lives[25]. Even in high resource settings like the US, racial and ethnic minorities have inadequate access to mental health services than whites and when they do receive care the quality is poor[23,24].

b.) *Mental health problems exist along a spectrum from mild, time-limited distress to chronic progressive and severely disabling conditions:* The underlying cause and mechanisms behind mental disorders are still an active area of research[26]. Scientists still do not fully understand the underlying cause or mechanism behind mental disorders which makes it difficult to manage what we cannot measure. To date, there is no objective test to measure the severity of mental health symptoms, and clinical diagnoses are routinely based on psychological evaluation using

retrospective self-reported and subjective expressed symptoms[27] that are known to be biased and inaccurate[28]. Other objective measures, such as MRI, are not useful for diagnosis as of yet, and even if they were, they are highly expensive and not widely available in rural and underserved communities.

Without an objective way to assess the behavioral symptoms and severity[29], it is possible for patients to have overlapping MHD symptoms and yet be diagnosed with only one mental health condition[30]. Some may be even missed all together if a symptom profiles don't fit Diagnostic and Statistical Manual of Mental Disorders[31] (DSM-5) criteria. For example, in order to be diagnosed with Major Depressive Disorder (MDD)[32] one of the most common types of mental disorders, five out of nine symptoms using the criteria listed in the DSM-5 (Table 1.1) should be met. The inherent heterogeneity of symptoms[33] and comorbidity[34] makes the psychiatric ailment classification particularly challenging. Moreover, present classification and distinction of MHD based on DSM-5 while useful from the nosological perspective may not adequately reflect the underlying dimensions, complexity, continuum, and severity of mental illness symptoms and the influence of external socio-environmental factors. In fact, the continued use of the DSM system of classification of mental illness is widely believed to be an impediment to new psychiatric research[35–38].

c.) *Mental health of individuals can be highly personalized and is influenced by a person's local social and environmental factors including genetic, neurodevelopmental, and psychological processes:* Psychological functioning can be effected through a complex interplay between inherited genetic traits[39,40] and additive effects of adverse life-events and socio-environmental factors[39,41] including urbanicity. In fact, the lack of robust genetic biomarkers for MHD could be

partly attributed to gene-environment interaction as the effect of genetic variants may only be seen in the presence of specific external environment-based stressors[32]. WHO report[42] emphasizes the need to understand social determinants of health to determine better preventative measures, as many of the preliminary causes and triggers of mental illness lie in social and economic spheres of daily life.

Table 1.1. DSM-5 criteria for major depressive disorder

"The individual must be **experiencing five or more symptoms during the same 2-week period** and at least one of the **symptoms should be either (1) depressed mood or (2) loss of interest or pleasure.** To receive a diagnosis of depression, these symptoms must cause the individual clinically significant distress or impairment in social, occupational, or other important areas of functioning. The symptoms must also not be a result of substance abuse or another medical condition"[43]
1. Depressed mood most of the day, nearly every day. 2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day. 3. Significant weight loss when not dieting or weight gain, or decrease or increase in appetite nearly every day. 4. A slowing down of thought and a reduction of physical movement (observable by others, not merely subjective feelings of restlessness or being slowed down). 5. Fatigue or loss of energy nearly every day. 6. Feelings of worthlessness or excessive or inappropriate guilt nearly every day. 7. Diminished ability to think or concentrate, or indecisiveness, nearly every day. 8. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

d.) *Mental Health is a fundamental human right and requires a systematic effort to reach out to the at-risk and marginalized population:* Although many countries have signed into law making it obligatory to provide mental health care to its citizens[44] the implementation of the law on the ground and an ability to track the outcomes objectively vary in both low[45,46] and high resource settings. 85% of the global population lives in low- and middle-income countries (LMIC) and over 400 million people do not have access to essential health care services[47]. Even where healthcare services are available the overall quality of care remains poor with noted disparities in access to the poor. For example in the US, despite several laws such as the Americans with Disabilities Act, the Individuals with Disabilities Education Act (IDEA), the Rehabilitation Services Act and more recently MHPAEA and ACA, significant gaps remain in providing minimally adequate mental health care to the at-risk and marginalized population[23,24]. With limited resources, the healthcare system fails to deliver timely, affordable and measurable care. Systematic reviews and meta-analyses indicate that behavioral interventions (e.g. cognitive behavioral therapy, problem-solving treatment) are effective in the treatment of depressive disorders[48–51]. Unfortunately, access to these treatments continues to be a problem[52], as only a fraction of individuals with mental illness, especially depression, use behavioral interventions[53,54]. These access-based problems are further exacerbated by poor engagement, as the modal number of in-person visits for behavioral interventions among depressed individuals is one[55,56].

In the last five years, there has been a noteworthy increase in the evaluation of technology[57–61] to help address some of the challenges highlighted above such as improving access, early remote assessment, deployment of behavioral interventions and personalization of treatment for improving the long term outcomes in MHD. There is also a growing discourse[62,63] between

researchers, clinicians, and technologists to build tools that can assess and detect the underlying triggers for behavioral health fluctuations at an early stage based on objective measures in addition to subjective assessments. The on-going research to transition reactive neuropsychiatry one that is based on clinical encounters that are episodic and rely on subjective assessments to a more proactive population level and remote MH monitoring system has been primarily driven by the growth and adoption of smartphones.

1.2 Smartphone-Enabled Solutions For Assessing And Intervening In Mental Health

In the last decade, mobile phones have enabled the last mile connectivity with 75% of the population living in LMIC (low and middle-income countries) owning a mobile phone and 64% of them having access to the internet[64]. The smartphone ownership has also steadily increased in LMICs[65] including minority and low-income populations[66,67]. In the developed economies, smartphone ownership[68] varies between 60-85%, with people spending as much as 150 minutes daily on their phones with >2,500 average screen touches[69–71]. This vast ubiquitous growing network of smart and connected devices[72] has enabled a cost-effective and scalable medium for researchers to reach and recruit participants for conducting remote biomedical health research[73] on a larger scale compared to traditional in-clinic research (Figure 1.1).

Mobile technology has the potential to reduce the complexity and cost of in-person clinical trials[74], by addressing challenges in the timely recruitment of a sufficiently large and diverse target population[75] as well as collecting in the moment data. The episodic in-person evaluations can often miss capturing important individual-specific and in the moment experiences of disease symptoms,

fluctuations and long-term disease progression (Figure 1.2) as it occurs outside the clinic. Mobile-based technologies, on the other hand, can help recruit a large and diverse sample significantly quickly and at a fraction of the cost. The technology also allows researchers an opportunity to collect and track momentary real-world data[76] from participants to help model disease symptoms and severity at an individual level along with the context of one's local environment. Rather than retrospectively asking people to recall their health over the past week or month, researchers using mobile technologies can assess participants functioning frequently and at important points in time (Figure 1.2) without having to wait until the next clinical visit and rely on recall that is known to have bias[77]. Rather than clinicians asking patients "how was your week" they can now say. "Let's review how your week was." The technology-enabled assessment of real-world experience at the population scale also presents a new paradigm for evaluating the efficacy of interventions in the real-world outside the controlled clinical settings. Such pragmatic clinical trials[78] are promising as they aim to evaluate the effects of an intervention under the usual real-world conditions using a diverse participant pool compared to "ideal circumstances" where traditional explanatory/randomized trials are conducted in.

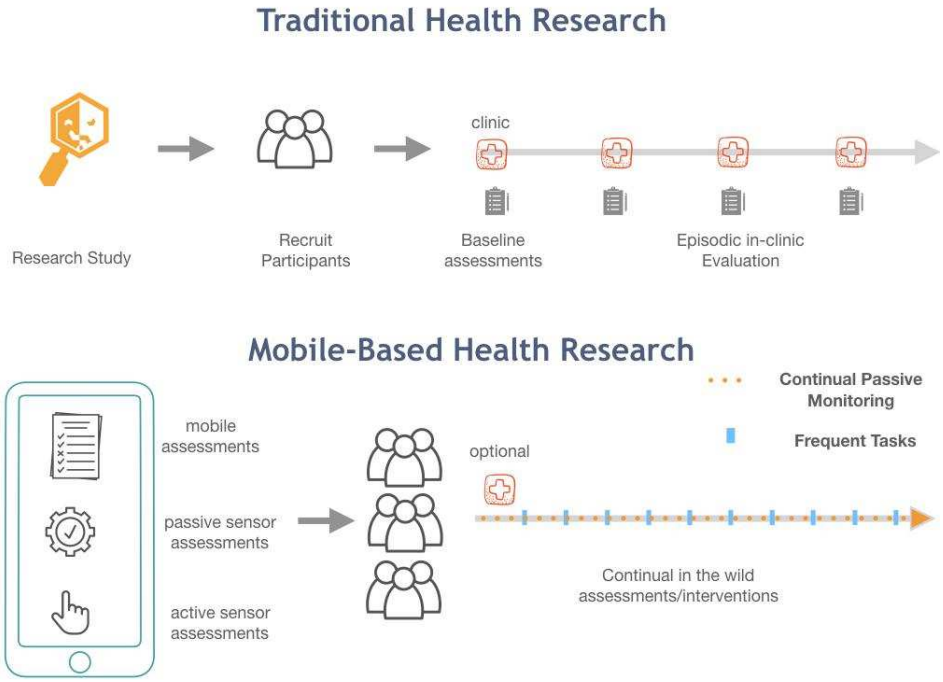


Figure 1.1. Comparison of a traditional health research model to a smartphone-mediated fully remote health research model.

The expansive mobile network also reaches the at-risk marginalized communities and population living in low resource settings. For example, 13% of Americans with annual income < $30,000 that solely rely on smartphones to address their internet needs[70]. This unique sub-population can be reached and assessed remotely only through smartphone-based health monitoring. With a wide reach and penetration, technology-aided platforms using smartphone apps may be well suited to fill in critical gaps in the current healthcare research model by providing remote health education, ambulatory assessment, and disease monitoring including the deployment of remote digital interventions in behavior health[79,80] to the last mile. Besides the ability to remotely monitor participants' health using subjective assessments (also referred to as patient/participant report

outcomes PRO's), the smartphones also enable the collection of high-frequency sensor-based data. These multi-faceted data streams obtained from smartphones often in a non-reactive manner (without active user input) could offer objective insights into people's local social and environmental factors that are well known to be linked to MHD[81,82]. The at-scale remote monitoring also enables tracking and potentially early detection of transition in depression states from normal to the symptomatic to help advance our understanding of the disease progression through real-world pragmatic evaluation. (Figure 1.2).

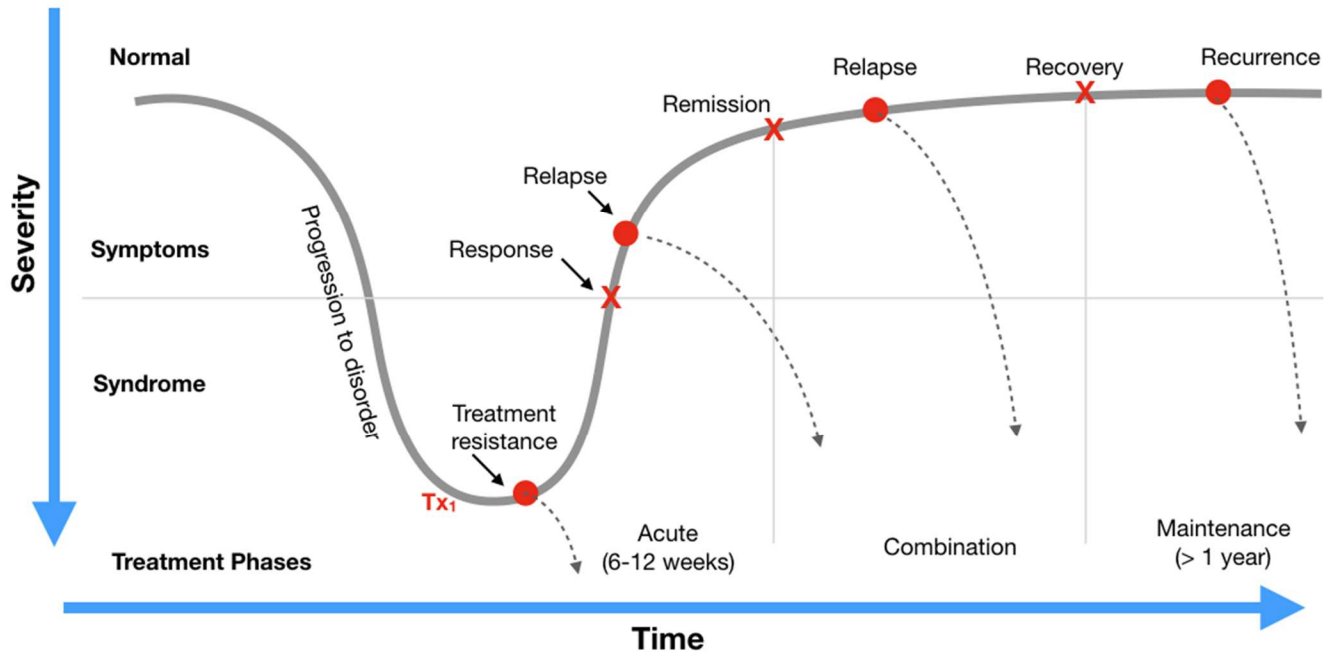


Figure 1.2. Phases of major depression treatment and its progression and management over time. Dashed lines indicate a potential worsening of depressive severity. Remission, the goal of treatment (pharmacological or psychotherapy), refers to the resolution of depressive symptoms and return to premorbid functioning; response refers to substantial clinical improvement which may or may not reach remission.

(Adapted from - https://www.ncbi.nlm.nih.gov/books/NBK338234/figure/introduction.f1/)

1.3 Utilizing Data Streams From Smartphones To Assess Behavioral Health

To date one of the major difficulties in transitioning from the current symptom-based assessment and classification of mental illness to one that is based on objective real-world evidence of behavior is due to the lack of robust and measurable quantitative features[35] to monitor behavioral health beyond the current clinical practice of subjective evaluations using survey instruments. To transform the psychiatric care and its ability to detect and diagnose mental conditions early and at scale, there is a critical need[83,84] to develop new ways to i) assess specific and underlying generative constructs of behavior patterns linked with neuropsychiatric conditions and ii) objectively quantify response to behavioral interventions. The real-world data[85] such as daily mobility, social interactions, etc generated from smartphone-based sensors, aggregated over time, could be specific constructs that are potentially indicative of one's behavioral health.

Smartphone technology has become increasingly sophisticated over time with an array of high fidelity onboard sensors[86]. With efficient battery utilization, the embedded phone-based sensors are able to continuously track a variety of human-smartphone interactions[87] including the ability to sense an individual's "life space[88]". Coupled with high-frequency daily device usage, a large volume of highly personalized "digital exhaust"[89] is being generated continuously. This multi-dimensional high-velocity data once processed and analyzed[90] could offer a wealth of semantic and contextual information to help build a personalized behavior profile that could be used to predict future behavior fluctuations. Combined together these longitudinal active and passive data streams offer an objective and systems-level approach for digital characterizations of human behavior. This analytical approach is broadly defined as *digital phenotyping* i.e. "moment-by-moment quantification of the individual-level human phenotype in situ using data from personal

digital devices"[91]. Typically sensor-based data is gathered either actively (requiring an active user input – e.g. finger-tapping task on-screen) or passively (requires no user input - eg. number of daily social interactions on phone). The latter is a non-reactive and less intrusive way (no burden on the participant) to collect real-world data and is often used in conjunction with self-reported subjective data.

The digital phenotyping in the present context of passive sensing is aimed at developing robust high-level behavioral markers[28,87,90,92–96] mapping the features derived from low-level sensors to clinical states (Figure 1.3). For example, GPS based location tracking could be used to generate low-level features such as overall physical activity and contextual features such as location (home, office) that could be indicative of hedonic behavior, fatigue, social avoidance etc. Similarly, the in-phone communication (#phones, #messages) could be suggestive of an individual's lack of social-activity and depressed mood. Besides the ability to detect early signs of behavioral aberrations, the low-burden passive tracking could help objectively quantify treatment response[97,98] for people undergoing psycho- or pharmaco-therapy. Identifying early resistance to interventions remains a significant challenge[99] especially in between the episodic clinical visits. Additionally, building a comprehensive mapping between passive data features and neuropsychiatric constructs could allow future studies to deploy objective and validated digital biomarkers as primary and secondary endpoints[100] in clinical trials. These can help further contextualize and trigger need-based subjective surveys and notifications to help assess response to behavioral interventions.

The sensor-based data also presents a unique opportunity to evaluate stratified patterns (sub-groups) of behavior variations based on "digital behavioral profiles" alone, independent of

subjective assessments. This unsupervised evaluation can help elucidate novel behavioral features from real-world data that may not be captured by subjective assessments and therefore may not even surface in objective (sensor-based)-subjective(survey-based) association analyses. The discovery of any such robust and sub-patterns based purely on the "digital data" has the potential to disrupt the current status quo of psychiatric ailment management based on episodic and subjective assessments only.

Several pilot studies in the last five years have examined the utility of assessing various mental disorders with many focusing on diagnosing depression[95,96,101–103,92,95,104–106]. While these early studies were able to demonstrate the utility of passive sensing in detecting depression severity, the findings at best showed a weak cohort level digital signature of depression based on GPS-derived mobility. The sample size was also relatively small (20-70 participants) and in most cases lacked diversity. Additionally, the analytical approaches did not compare the role of participant demographics on depression prediction relative to passive data. As also suggested by groups that conducted these early research studies, further work is needed to both replicate their findings in a large nationally recruited and diverse cohort. The density and longitudinal nature of collected data at an individual level also offer the opportunity to assess the potential of truly "N-of-1" precision psychiatry. Data from each participant can be used to create an individualized baseline profile "average digital behavior", the deviations from which could be used to predict behavior anomalies using personalized machine learning models.

However, the promise of such data-driven efforts is dependent on people's willingness to participate and sustainably engage in remote online studies through smartphone-based apps and share their digital data with researchers. Early trends have shown engagement in remote research

to be particularly challenging[107] highlighting an urgent need for further research to assess trends in participant engagement using quantitative(empirical evaluation) and qualitative(evaluate the reason behind empirical findings) methods.

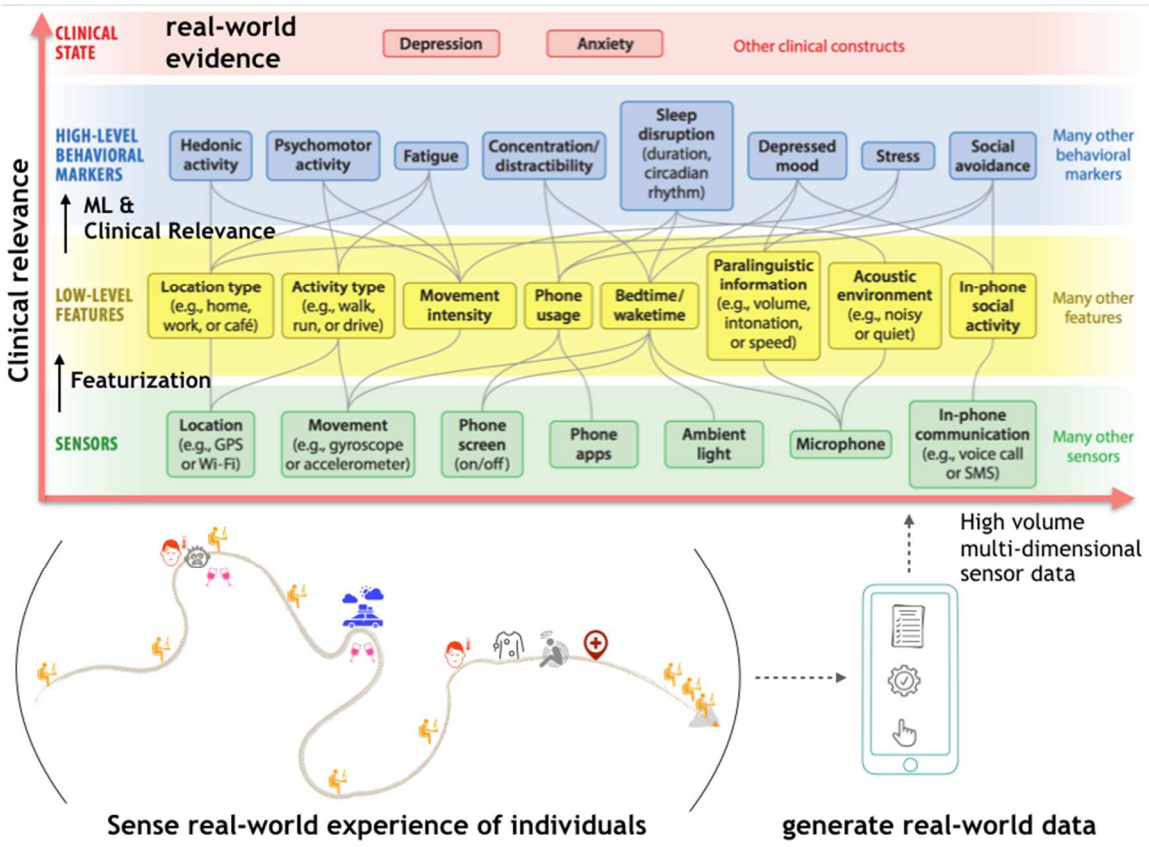


Figure 1.3. Overall schematic of smartphone-based sensing of behavioral health.
The multi-dimensional raw data captured from sensors can be featurized to generate low-level features such as daily mobility, phone usage, etc which can then be utilized by machine learning models to generate high-level behavioral markers that can be further selected working together with domain clinical experts.

Part of this figure has been adapted from Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning | David C. Mohr, Mi Zhang, Stephen M. Schueller | Annual Review of Clinical Psychology 2017 13:1, 23-47

1.4 Understanding People's Willingness To Accept And Utilize Technology For Biomedical Research

Over the last five years, there has been significant growth in the use of various online mediums such as social media, web and smartphones-based apps for reaching and enrolling a large number of study participants in online biomedical research. While these "fully remote" studies have shown promising results in enrolling and collecting real-world data from large cohorts, essentially disrupting traditional in-clinic research; they have also surfaced pertinent challenges in participant engagement[107]. In fact, the present situation is no different than the internet-based trials[108,109] in the early 2000s where high user attrition surfaced as significant challenge[110–112]. However, our understanding of participant attrition in remote research continues to be limited with a lack of evidence-based guidance on participant recruitment and retention strategies for digital health. One reason why "user attrition" continues to be poorly understood is due to the focus of researchers in discovering the clinically relevant insights from the collected real-world data. Often times, a smaller sub-cohort of participants that contributed the largest amount of longitudinal data (most engaged) in the remote study is selected for analysis purposes. Very few studies publish and share insights on potential differences in participant characteristics across the cohort chosen for analysis compared to participants who were dropped out. While it may be feasible for early exploratory analysis, the selection of participants based on their engagement in the study can introduce severe selection and ascertainment bias which can impact the validity and generalizability of the findings such studies[113]. Furthermore, with the analysis focus only on people who contributed sufficient data, we lose an opportunity to systematically evaluate any significantly differential patterns between people "who" remained engaged in the study for longer duration vs people who left early within a day or two of joining the study. The existing real-world data collected in the past digital

health studies can be further mined to identify potential patterns in the participant enrollment, retention, and long term-app usage. Any significant findings from a large-scale cross-study user enrollment and retention analysis can help inform the development and design of future digital studies in the real-world

Finally, the highly personalized data collected from smartphones especially in the context of mental health also raises data privacy and misuse concerns for the study participants. Several large-scale data privacy violations reported in the last 12 months[114–116] and researchers expressing concerns about data sharing policies[117,118] of mental health apps, can further impact people's continued willingness to join large-scale online remote research and share personal digital data for research. However, there is no contemporary evidence available on people's willingness to join an online study and share their data varies based on the institution (academic, federal or pharma) of researchers conducting the study, the recruitment platform (Google Vs Facebook).

1.5 Specific Aims And Dissertation Outline

The overall aim of my research is to investigate the potential of assessing and mediating in mental health using digital platforms fully remotely and evaluate the utility of generated data in predicting mental health outcomes. Additionally, and equally importantly I evaluate participants' retention in online digital health research studies as well as their future willingness to participate and share digital data in online biomedical research. Specifically, I have focused on four research topics (Aims 1-4), the results from which are aimed to help improve our understanding of utilizing technology-based solutions to assess and intervene in mental health.

AIM 1: Evaluate the utility of active and passive data collected fully remotely through smartphones for assessing symptoms of major depressive disorder (MDD) at the population and

individual level. Additionally, assess the presence of any underlying substructure of depression symptomatology in the collected real-world evidence. (Chapter 2)

AIM 2: Evaluate the feasibility of deploying digital tools for conducting research studies fully remotely in a limited resource setting with minorities such as Hispanic/Latinos to ascertain how they interact with mHealth apps, and the potential clinical impact apps may have on treating depression in this minority population. (Chapter 3)

AIM 3: Assess participant enrollment and retention in large-scale digital health research studies to ascertain underlying trends in technology access and utilization by different sub-groups based on socioeconomic status, race/ethnicity, age, gender and incentives offered. (Chapter 4)

AIM 4: Evaluate individual's willingness to participate and share their digital data in online biomedical research (Chapter 5).

1.6 RELEVANCE

These research aims are also contemporary in nature and aligned with the strategic goals of various national and international agencies working to improve mental health outcomes. In particular NIMH Strategic Research Priorities[119], Section 2.2 *"…identify, early in the development of major mental illnesses, biomarkers and behavioral indicators with high predictive value to guide the use of preventive interventions. …to develop biomarkers and assessment tools to predict illness onset, course, and intervention response across diverse populations…for stratification purposes, (to)…."* and also the technology-based opportunities highlighted in the 2017 NIMH Council report[120]. These aims also help address some of the critical challenges highlighted in 2018 Lancet report (Chapter-1. a-d) and overlaps with IOM's triple aim by evaluating utility of remote technology-based tools that could be deployed at scale, increasing access cutting across socioeconomic and

cultural barriers, and offer citizens alternatives to assess and track their individualistic lived experience of disease in an objective, momentary and nonreactive ways[90,93,121].

1.7 REFERENCES

1. Kraepelin, E. Die Erscheinungsformen des Irreseins. *History of Psychiatry* **3**, 509–529 (1992).

2. Wang, P. S., Demler, O. & Kessler, R. C. Adequacy of Treatment for Serious Mental Illness in the United States. *Am. J. Public Health* **92**, 92–98 (2002).

3. Steel, Z. *et al.* The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. *Int. J. Epidemiol.* **43**, 476–493 (2014).

4. Agency for Healthcare Research & Quality. Medical Expenditure Panel Survey Publication Details. Available at: https://meps.ahrq.gov/data_stats/Pub_ProdResults_Details.jsp?pt=Statistical+Brief&opt=2&id=910. (Accessed: 27th October 2018)

5. Roehrig, C. Mental Disorders Top The List Of The Most Costly Conditions In The United States: $201 Billion. *Health Aff.* 10.1377/hlthaff.2015.1659 (2016).

6. [No title]. Available at: https://store.samhsa.gov/shin/content/SMA14-4883/SMA14-4883.pdf. (Accessed: 4th February 2018)

7. Treatment prevalence - Peterson-Kaiser Health System Tracker. *Peterson-Kaiser Health System Tracker* Available at: https://www.healthsystemtracker.org/indicator/spending/treatment-prevalence/. (Accessed: 26th October 2018)

8. GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* **388**, 1603–1658 (2016).

9. What are the current costs and outcomes related to mental health and substance abuse disorders? - Peterson-Kaiser Health System Tracker. *Peterson-Kaiser Health System Tracker* Available at: https://www.healthsystemtracker.org/chart-collection/current-costs-outcomes-related-mental-health-substance-abuse-disorders/. (Accessed: 27th October 2018)

10. Website. Available at: https://fas.org/sgp/crs/misc/R43047.pdf. (Accessed: 25th October 2018)

11. The State of Mental Health in America. *Mental Health America* (2015). Available at: http://www.mentalhealthamerica.net/issues/state-mental-health-america. (Accessed: 25th October 2018)

12. Fact Sheet Library | NAMI: National Alliance on Mental Illness. Available at: https://www.nami.org/Learn-More/Fact-Sheet-Library. (Accessed: 25th October 2018)

13. Murray, C. J. L. *et al.* The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA* **310**, 591–608 (2013).

14. NIMH » Mental Illness. Available at: https://www.nimh.nih.gov/health/statistics/mental-illness.shtml. (Accessed: 6th March 2018)

15. WHO | Mental health action plan 2013 - 2020. (2015).

16. WHO | Sustainable Development Goals (SDGs). (2017).

17. Patel, V. *et al.* The Lancet Commission on global mental health and sustainable development. *Lancet* (2018). doi:10.1016/S0140-6736(18)31612-X

18. Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).

19. WHO | Suicide rates (per 100 000 population). (2018).

20. [No title]. Available at: https://www.nature.com/polopoly_fs/1.19694!/menu/main/topColumns/topLeftColumn/pdf/532020a.pdf?origin=ppub. (Accessed: 25th October 2018)

21. Baylor, C. Mental Health Treatment EBP | SAMHSA - Substance Abuse and Mental Health Services Administration. Available at: https://www.samhsa.gov/ebp-web-guide/mental-health-treatment. (Accessed: 25th October 2018)

22. Lund, C. Improving quality of mental health care in low-resource settings: lessons from PRIME. *World Psychiatry* **17**, 47–48 (2018).

23. US Department of Health and Human Services; Substance Abuse and Mental Health Services Administration. Mental Health: Culture, Race, and Ethnicity: A sment to Mental Health A Report of the Surgeon General. *PsycEXTRA Dataset* (2001). doi:10.1037/e415842005-001

24. Bussing, R. & Gary, F. A. Eliminating Mental Health Disparities by 2020: Everyone's Actions Matter. *J. Am. Acad. Child Adolesc. Psychiatry* **51**, 663–666 (2012).

25. WHO | World Health Organization. Available at: http://gamapserver.who.int/gho/interactive_charts/mental_health/psychiatrists_nurses/atlas.html. (Accessed: 21st October 2019)

26. NIMH » Division of Neuroscience and Basic Behavioral Science (DNBBS). Available at: https://www.nimh.nih.gov/about/organization/dnbbs/index.shtml. (Accessed: 4th November 2018)

27. Odp, O. 12.00-Mental Disorders-Adult. (2003).

28. Glenn, T. & Monteith, S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. *Curr. Psychiatry Rep.* **16**, 523 (2014).

29. Schmidt, H. D., Shelton, R. C. & Duman, R. S. Functional Biomarkers of Depression: Diagnosis, Treatment and Pathophysiology. *Neuropsychopharmacology* **36**, 2375–2394 (2011).

30. Franklin, J. C. *et al.* Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychol. Bull.* **143**, 187–232 (2017).

31. Psychiatry Online | DSM Library. Available at: https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596. (Accessed: 27th October 2018)

32. Otte, C. *et al.* Major depressive disorder. *Nature Reviews Disease Primers* **2**, 16065 (2016).

33. Website. Available at: http://psycnet.apa.org/record/1995-97231-011. (Accessed: 3rd November 2018)

34. Buckholtz, J. W. & Meyer-Lindenberg, A. Psychopathology and the Human Connectome: Toward a Transdiagnostic Model of Risk For Mental Illness. *Neuron* **74**, 990–1004 (2012).

35. Wiecki, T. V., Poland, J. & Frank, M. J. Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry. *Clin. Psychol. Sci.* **3**, 378–399 (2015).

36. Hyman, S. E. Revolution Stalled. *Sci. Transl. Med.* **4**, 155cm11–155cm11 (2012).

37. Insel, T. *et al.* Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).

38. Wardenaar, K. J. & de Jonge, P. Diagnostic heterogeneity in psychiatry: towards an empirical solution. *BMC Med.* **11**, 201 (2013).

39. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).

40. Lohoff, F. W. Overview of the Genetics of Major Depressive Disorder. *Curr. Psychiatry Rep.* **12**, 539–546 (2010).

41. Guze, S. B. Biological psychiatry: is there any other kind?*. *Psychol. Med.* **19**, 315 (1989).

42. [No title]. Available at: http://apps.who.int/iris/bitstream/handle/10665/112828/9789241506809_eng.pdf?sequence=1. (Accessed: 27th October 2018)

43. Depression Definition and DSM-5 Diagnostic Criteria. *Psycom.net - Mental Health Treatment Resource Since 1986* Available at: https://www.psycom.net/depression-definition-dsm-5-diagnostic-criteria/. (Accessed: 22nd October 2019)

44. WHO | WHO Country Profiles: Mental Health in Development (WHO proMIND). (2015).

45. Doku, V. C. K., Wusu-Takyi, A. & Awakame, J. Implementing the Mental Health Act in Ghana: any challenges ahead? *Ghana Med. J.* **46**, 241–250 (2012).

46. Wayne Holden, E. & Brannan, A. M. *Evaluating Systems of Care: The Comprehensive Community Mental Health Services for Children and Their Families Program. A Special Issue of Children's Services: Social Policy, Research, and Practice*. (Psychology Press, 2014).

47. WHO | New report shows that 400 million do not have access to essential health services. (2015).

48. Cuijpers, P. *et al.* Psychological treatment of depression in inpatients: a systematic review and meta-analysis. *Clin. Psychol. Rev.* **31**, 353–360 (2011).

49. Cuijpers, P. *et al.* Interpersonal psychotherapy for depression: a meta-analysis. *Am. J. Psychiatry* **168**, 581–592 (2011).

50. Driessen, E. *et al.* The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis update. *Clin. Psychol. Rev.* **42**, 1–15 (2015).

51. Cuijpers, P., Donker, T., van Straten, A., Li, J. & Andersson, G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychol. Med.* **40**, 1943–1957 (2010).

52. Collins, K. A., Westra, H. A., Dozois, D. J. A. & Burns, D. D. Gaps in accessing treatment for anxiety and depression: challenges for the delivery of care. *Clin. Psychol. Rev.* **24**, 583–616 (2004).

53. Harpaz-Rotem, I., Libby, D. & Rosenheck, R. A. Psychotherapy use in a privately insured population of patients diagnosed with a mental disorder. *Soc. Psychiatry Psychiatr. Epidemiol.* **47**, 1837–1844 (2012).

54. Marcus, S. C. & Olfson, M. National trends in the treatment for depression from 1998 to 2007. *Arch. Gen. Psychiatry* **67**, 1265–1273 (2010).

55. Simon, G. E. & Ludman, E. J. Predictors of early dropout from psychotherapy for depression in community practice. *Psychiatr. Serv.* **61**, 684–689 (2010).

56. Olfson, M. *et al.* National trends in the outpatient treatment of depression. *JAMA* **287**, 203–209 (2002).

57. Anguera, J. A., Jordan, J. T., Castaneda, D., Gazzaley, A. & Areán, P. A. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. *BMJ Innov* **2**, 14–21 (2016).

58. Glick, G., Druss, B., Pina, J., Lally, C. & Conde, M. Use of mobile technology in a community mental health setting. *J. Telemed. Telecare* **22**, 430–435 (2016).

59. Whittaker, R., McRobbie, H., Bullen, C., Rodgers, A. & Gu, Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst. Rev.* **4**, CD006611 (2016).

60. Widmer, R. J. *et al.* Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. *Mayo Clin. Proc.* **90**, 469–480 (2015).

61. Maulik, P. K. *et al.* Systematic Medical Appraisal, Referral and Treatment (SMART) Mental Health Programme for providing innovative mental health care in rural communities in India. *Glob Ment Health (Camb)* **2**, e13 (2015).

62. Insel, T. R. Digital Phenotyping: Technology for a New Science of Behavior. *JAMA* **318**, 1215–1216 (2017).

63. Chauvin, J. J. & Insel, T. R. Building the Thermometer for Mental Health. *Cerebrum* **2018**, (2018).

64. Poushter, J., Bishop, C. & Chwe, H. Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones. *Pew Research Center's Global Attitudes Project* (2018). Available at: http://www.pewglobal.org/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/. (Accessed: 22nd July 2018)

65. Social Media Use Continues to Rise in Developing Countries. *Pew Research Center's Global Attitudes Project* (2018). Available at: http://www.pewglobal.org/2018/06/19/2-smartphone-ownership-on-the-rise-in-emerging-economies/. (Accessed: 28th October 2018)

66. Lewis, T., Synowiec, C., Lagomarsino, G. & Schweitzer, J. E-health in low- and middle-income countries: findings from the Center for Health Market Innovations. *Bull. World Health Organ.* **90**, 332–340 (2012).

67. Quintanilla, E. Cellphones helping minorities close gap on Internet access? *The Christian Science Monitor* (2011). Available at: https://www.csmonitor.com/USA/Society/2011/0210/Cellphones-helping-minorities-close-gap-on-Internet-access. (Accessed: 28th October 2018)

68. Top Countries/Markets by Smartphone Penetration & Users | Newzoo. *Newzoo* Available at: https://newzoo.com/insights/rankings/top-50-countries-by-smartphone-penetration-and-users/. (Accessed: 28th October 2018)

69. Request, T. How Much Time Do People Spend on Their Mobile Phones in 2017? *Hacker Noon* (2017). Available at: https://hackernoon.com/how-much-time-do-people-spend-on-their-mobile-phones-in-2017-e5f90a0b10a6. (Accessed: 16th October 2017)

70. Website. Available at: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/. (Accessed: 16th October 2017)

71. Winnick, M. Putting a Finger on Our Phone Obsession. Available at: https://blog.dscout.com/mobile-touches. (Accessed: 16th October 2017)

72. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. *Pew Research Center's Global Attitudes Project* (2019). Available at: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/. (Accessed: 30th July 2019)

73. Steinhubl, S. R. & Topol, E. J. Digital medicine, on its way to being just plain medicine. *NPJ Digit Med* **1**, 20175 (2018).

74. Marquis-Gravel, G. *et al.* Technology-Enabled Clinical Trials: Transforming Medical Evidence Generation. *Circulation* **140**, 1426–1436 (2019).

75. Fogel, D. B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp Clin Trials Commun* **11**, 156–164 (2018).

76. ElZarrad, M. K., Khair ElZarrad, M. & Corrigan□Curay, J. The US Food and Drug Administration's Real□World Evidence Framework: A Commitment for Engagement and Transparency on Real□World Evidence. *Clinical Pharmacology & Therapeutics* (2019). doi:10.1002/cpt.1389

77. Althubaiti, A. Information bias in health research: definition, pitfalls, and adjustment methods. *J. Multidiscip. Healthc.* **9**, 211–217 (2016).

78. Tuzzio, L. & Larson, E. B. The Promise of Pragmatic Clinical Trials Embedded in Learning Health Systems. *EGEMS (Wash DC)* **7**, 10 (2019).

79. Patel, V., Mishra, P. & Patni, J. C. PsyHeal: An Approach to Remote Mental Health

Monitoring System. in *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (2018). doi:10.1109/icacce.2018.8441760

80. Mohr, D. C., Burns, M. N., Schueller, S. M., Clarke, G. & Klinkman, M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *Gen. Hosp. Psychiatry* **35**, 332–338 (2013).

81. Daniel, H., Bornstein, S. S., Kane, G. C. & Health and Public Policy Committee of the American College of Physicians. Addressing Social Determinants to Improve Patient Care and Promote Health Equity: An American College of Physicians Position Paper. *Ann. Intern. Med.* **168**, 577–578 (2018).

82. Allen, J., Balfour, R., Bell, R. & Marmot, M. Social determinants of mental health. *Int. Rev. Psychiatry* **26**, 392–407 (2014).

83. Anticevic, A. & Murray, J. D. *Computational Psychiatry: Mathematical Modeling of Mental Illness*. (Academic Press, 2017).

84. Ferrante, M. *et al.* Computational psychiatry: a report from the 2017 NIMH workshop on opportunities and challenges. *Mol. Psychiatry* (2018). doi:10.1038/s41380-018-0063-z

85. Jarow, J. P., LaVange, L. & Woodcock, J. Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using 'Real-World' Data. *JAMA* **318**, 703–704 (2017).

86. Priyadarshini, M. Which Sensors Do I Have In My Smartphone? How Do They Work? *Fossbytes* (2018). Available at: https://fossbytes.com/which-smartphone-sensors-how-work/. (Accessed: 22nd October 2019)

87. Cornet, V. P. & Holden, R. J. Systematic review of smartphone-based passive sensing for health and wellbeing. *J. Biomed. Inform.* **77**, 120–132 (2018).

88. Wettstein, M., Wahl, H.-W. & Schwenk, M. Life Space in Older Adults. in *Oxford Research Encyclopedia of Psychology* (Oxford University Press, 2018).

89. Insel, T. R. Digital phenotyping: a global tool for psychiatry. *World Psychiatry* **17**, 276–277 (2018).

90. Mohr, D. C., Zhang, M. & Schueller, S. M. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annu. Rev. Clin. Psychol.* **13**, 23–47 (2017).

91. Torous, J., Kiang, M. V., Lorme, J. & Onnela, J.-P. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Ment Health* **3**, e16 (2016).

92. Servia-Rodríguez, S. *et al.* Mobile Sensing at the Service of Mental Well-being. in *Proceedings of the 26th International Conference on World Wide Web - WWW '17* (2017). doi:10.1145/3038912.3052618

93. Onnela, J.-P. & Rauch, S. L. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology* **41**, 1691–1696 (2016).

94. Goodspeed, R. *et al.* Comparing the Data Quality of Global Positioning System Devices and Mobile Phones for Assessing Relationships Between Place, Mobility, and Health: Field Study. *JMIR Mhealth Uhealth* **6**, e168 (2018).

95. Saeb, S. *et al.* Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J. Med. Internet Res.* **17**, e175 (2015).

96. Batterham, P. Peer Review #1 of 'The relationship between mobile phone location sensor data and depressive symptom severity (v0.1)'. (2016). doi:10.7287/peerj.2537v0.1/reviews/1

97. Fava, M. Diagnosis and definition of treatment-resistant depression. *Biol. Psychiatry* **53**, 649–659 (2003).

98. Souery, D. *et al.* Treatment resistant depression: methodological overview and operational criteria. *Eur. Neuropsychopharmacol.* **9**, 83–91 (1999).

99. Knöchel, C. *et al.* Treatment-resistant Late-life Depression: Challenges and Perspectives. *Curr. Neuropharmacol.* **13**, 577–591 (2015).

100. Clinical Endpoint - an overview | ScienceDirect Topics. Available at: https://www.sciencedirect.com/topics/medicine-and-dentistry/clinical-endpoint. (Accessed: 20th October 2019)

101. Madan, A., Cebrian, M., Lazer, D. & Pentland, A. Social sensing for epidemiological behavior change. *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10* (2010). doi:10.1145/1864349.1864394

102. Wang, R. *et al.* StudentLife: Using Smartphones to Assess Mental Health and Academic Performance of College Students. *Mobile Health* 7–33 (2017). doi:10.1007/978-3-319-51394-2_2

103. Canzian, L. & Musolesi, M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15* 1293–1304 (ACM Press, 2015).

104. Burns, M. N. *et al.* Harnessing context sensing to develop a mobile intervention for

depression. *J. Med. Internet Res.* **13**, e55 (2011).

105. Place, S. *et al.* Behavioral Indicators on a Mobile Sensing Platform Predict Clinically Validated Psychiatric Symptoms of Mood and Anxiety Disorders. *J. Med. Internet Res.* **19**, e75 (2017).

106. StudentLife Study. Available at: http://studentlife.cs.dartmouth.edu/. (Accessed: 29th October 2018)

107. Druce, K. L., Dixon, W. G. & McBeth, J. Maximizing Engagement in Mobile Health Studies: Lessons Learned and Future Directions. *Rheum. Dis. Clin. North Am.* **45**, 159–172 (2019).

108. Christensen, H., Griffiths, K. M., Korten, A. E., Brittliffe, K. & Groves, C. A Comparison of Changes in Anxiety and Depression Symptoms of Spontaneous Users and Trial Participants of a Cognitive Behavior Therapy Website. *Journal of Medical Internet Research* **6**, e46 (2004).

109. Christensen, H., Griffiths, K. M. & Jorm, A. F. Delivering interventions for depression by using the internet: randomised controlled trial. *BMJ* **328**, 265 (2004).

110. Eysenbach, G. The Law of Attrition. *Journal of Medical Internet Research* **7**, e11 (2005).

111. Christensen, H. & Mackinnon, A. The law of attrition revisited. *Journal of medical Internet research* **8**, e20; author reply e21 (2006).

112. Eysenbach, G. The Law of Attrition Revisited – Author's Reply. *Journal of Medical Internet Research* **8**, e21 (2006).

113. Kaplan, R. M., Chambers, D. A. & Glasgow, R. E. Big data and large sample size: a cautionary note on the potential for bias. *Clin. Transl. Sci.* **7**, 342–346 (2014).

114. US CMS says 75,000 individuals' files accessed in data breach. *Deccan Chronicle* (2018). Available at: https://www.deccanchronicle.com/technology/in-other-news/201018/us-cms-says-75000-individuals-files-accessed-in-data-breach.html. (Accessed: 16th September 2019)

115. Perez, S. & Whittaker, Z. Everything you need to know about Facebook's data breach affecting 50M users. *TechCrunch* (2018). Available at: http://social.techcrunch.com/2018/09/28/everything-you-need-to-know-about-facebooks-data-breach-affecting-50m-users/. (Accessed: 16th September 2019)

116. Wakabayashi, D. Google Plus Will Be Shut Down After User Information Was Exposed. (2018). Available at: https://www.nytimes.com/2018/10/08/technology/google-plus-security-disclosure.html. (Accessed: 16th September 2019)

117. Huckvale, K., Torous, J. & Larsen, M. E. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. *JAMA Netw Open* **2**, e192542 (2019).

118. Torous, J. & Roberts, L. W. Needed Innovation in Digital Health and Smartphone Applications for Mental Health: Transparency and Trust. *JAMA Psychiatry* **74**, 437–438 (2017).

119. NIMH » Strategic Research Priorities Overview. Available at: https://www.nimh.nih.gov/about/strategic-planning-reports/strategic-research-priorities/index.shtml. (Accessed: 25th October 2018)

120. NIMH » Opportunities and Challenges of Developing Information Technologies on Behavioral and Social Science Clinical Research. Available at: https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/opportunities-and-challenges-of-developing-information-technologies-on-behavioral-and-social-science-clinical-research.shtml. (Accessed: 29th October 2018)

121. NIMH » Opportunities and Challenges of Developing Information Technologies on Behavioral and Social Science Clinical Research. Available at: https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/opportunities-and-challenges-of-developing-information-technologies-on-behavioral-and-social-science-clinical-research.shtml. (Accessed: 28th October 2018)

Chapter 2. THE ACCURACY OF PASSIVE PHONE SENSORS IN PREDICTING DAILY MOOD

2.1 ABSTRACT

Background:

Smartphones provide a low-cost and efficient means to collect population level data. Several small studies have shown promise in predicting mood variability from smartphone-based sensor and usage data, but have not been generalized to nationally recruited samples. This study used passive smartphone data, demographic characteristics, and baseline depressive symptoms to predict prospective daily mood.

Method:

Daily phone usage data was collected passively from 271 Android phone users participating in a fully remote randomized controlled trial of depression treatment (BRIGHTEN). Participants completed daily PHQ-2 questionnaires. A machine learning approach was used to predict daily mood for the entire sample and individual participants.

Results:

Sample-wide estimates showed a marginally significant association between physical mobility and self-reported daily mood (B = -0.04, $p < .05$), but the predictive models performed poorly for the sample as a whole (median $R^2 \sim 0$). Focusing on individuals, 13.9% of participants showed significant association (*FDR < .10*) between a passive feature and daily mood. Personalized models combining features provided better prediction performance (median AUC > .50) for 80.6% of participants and very strong prediction in a subset (median AUC > .80) for 11.8% of participants.

Conclusions:

Passive smartphone data with current features may not be suited for predicting daily mood at a population level because of the high degree of intra- and inter-individual variation in phone usage patterns and daily mood ratings. Personalized models show encouraging early signs for predicting an individual's mood state changes, with GPS-derived mobility being the top most important feature in the present sample.

2.2 INTRODUCTION

Depressive disorders are among the leading causes of disability and mortality globally 1. Although effective depression treatments exist 2, the sequelae of depressive disorders continue to rise: 10 years ago, depression was the 5th leading cause of morbidity; now, it is the leading cause (World Health Organization, 20123. One factor complicating the detection and treatment of depression is the use of sporadically collected self-report assessments. Although validated measures like the PHQ-9 are useful tools for measurement-based care, they only reflect perceived mood over the past two weeks, which is subject to temporal bias, and they typically only assess mood symptoms, not functional symptoms 4. Health care organizations and clinicians face an additional challenge when patients fail to return for appointments: Is this because their condition has worsened or because it has improved substantially and there is no need for further treatment? As one recent study 5 found, both scenarios are true: some patients do not return because they are not responding to treatment and their condition is worsening; others do not return because they no longer have the need.

A partial solution to these problems is ecological momentary assessment (EMA)6, which may leverage smartphone data to enhance clinical decision making for depression. By collecting

information about mood and function as it occurs in real time, EMA captures continuous data regarding symptoms and behavior, which can create a more accurate and complete picture of treatment response. Mobile technology can serve as an acceptable, low-cost, and efficient means of collecting this information. These technologies have long supported active data capture, such as in the form of smartphone-based questionnaires, but in recent years, mobile health (mHealth) developers have turned to passive data collection via the use of device sensors, information from online calendars, and number of people contacted via telecommunication technologies 7,8. Use of text messages and email may serve as a proxy for engagement and social connectedness 9, an important measure of functioning and treatment response in depression. Several small studies have found preliminary evidence that activity based on smartphone global positioning system (GPS) and accelerometry can predict depressed mood 10–12. However, most recently, 13 found weak and inconsistent relationships between specific location data derived from location data (e.g., work, home, shopping, place of worship) and symptoms of depression and anxiety. Thus, there is ongoing uncertainty regarding mobility and GPS data as predictors of mental health.

It is critical that studies of the predictive capacity of passive data move beyond small, homogeneous samples to better characterize the true potential of such assessment in the population as a whole. Our previous work demonstrated the feasibility and cost effectiveness of a large, fully remote randomized controlled trial (RCT) of depression intervention 14. This secondary analysis of the BRIGHTEN study examined whether features of typical smartphone usage (e.g., texts, calls) and sensor data (e.g., mobility based on GPS) predicted mood beyond the variance explained by demographics and baseline depressive symptoms. We used machine learning to predict future self-reported daily mood from passive data both within the entire sample and systematically examined

interindividual heterogeneity using personalized N-of-1 models for predicting an individual's daily mood.

2.3 Materials and Methods

2.3.1 *Participants*

Ethical approval for the BRIGHTEN study was given by the UCSF Committee for Human Research. Participants were recruited across all 50 U.S. states via Craigslist, Google AdWords™, and Twitter™, as well as shuttle advertisements in the San Francisco Bay Area. Eligible participants were 18 years or older, able to read English, had a smartphone (Android or iPhone) with WiFi or 3G/4G capabilities, and obtained a score of five or more on the Patient Health Questionnaire-9 (PHQ-9[15], and/or indicated that their depressive symptoms made it "very" or "extremely" difficult to function at work, home, or socially.

2.3.2 *Procedures*

A full description of the procedures for the BRIGHTEN can be found in [14]. Briefly, BRIGHTEN was a large, fully remote RCT of depression treatment. Interested participants were directed to an online portal where they watched an informational video describing the study and provided informed consent. Eligible participants were randomized to one of three apps. Treatment and assessment for the parent trial was delivered via participants' smartphones. In addition to completing a demographics questionnaire and baseline PHQ-9, participants reported daily mood through an assessment app, and passive data was captured through Ginger.io app™. Participants engaged in treatment for the first month of the study, and continued follow-up assessments for two months post-treatment. Participants were paid $20 for each assessment at 4, 8, and 12 weeks.

2.3.3 *Measures*

Participants were prompted to complete a daily two-item Patient Health Questionnaire (PHQ-2)[16] assessing depressive symptoms of mood ("Feeling down, depressed, or hopeless") and anhedonia ("Little interest or pleasure in doing things"). The PHQ-2 items were modified to inquire about symptoms over the past 24 hours using a modified 5-point rating scale (1 = *not at all*; 5 = *most of the day*; possible scores ranging from 2-10). Participants gave permission to have some measures of typical phone usage collected passively (i.e., collected in the background without user involvement). From the collected raw phone usage and sensor data, passive features (see Table 2.1) were generated by Ginger.io. Phone-based variables were aggregated into 24-hour periods. For each passive feature, we also computed the daily deviation from an individual's median value of that feature.

2.3.4 *Data analyses*

Prior to the analysis, any missing passive or self-reported mood data were imputed using a participant's median weekly value per feature. PHQ-2 scores were aligned to the passive data so that they referred to the same 24-hour period. We used generalized estimating equations (GEEs) [17] to assess the marginal association between longitudinal daily mood and passive phone data in the sample. GEE models extend generalized linear models to longitudinal or clustered data using a working correlation structure that accounts for within-subject correlations of daily responses, thereby estimating robust and unbiased standard errors compared to ordinary least squares regression [17,18]. For machine learning analyses predicting daily mood from phone-based features, we used an ensemble-based method called random forests [19], which show robust and strong prediction across many types of data, particularly in the biomedical domain [20,21]. A random forest

model bootstraps many versions of the data via sampling with replacement, and then on each new dataset, the model fits a shallow decision tree, which is an alternative form of regression that allows nonlinear associations and complex interactions. It is an ensemble method because the decision tree models across many bootstrapped datasets are combined into a final prediction model. In our analyses, three classes of predictors were included in the models: a) baseline demographics (gender, age, marital status, and race/ethnicity), b) baseline PHQ-9 score, and c) daily phone usage features. These predictor classes were added sequentially across three models.

We predicted daily PHQ-2 score for both the whole sample and each individual. In each case, models were trained to predict a person's daily mood based on the passive data from the previous 24 hours' phone usage and the available demographic variables for the cohort. The primary statistic of interest for the marginal model was R^2, assessing how close the model predictions are to the true values in the test data. Given that the PHQ-2 response scale was modified for daily responding in this study, there are no established clinical cutoffs. Therefore, for these exploratory person-specific models, we predicted two discrete mood state groups: those with no symptomatology (PHQ-2 = 2) and those reporting symptoms (PHQ-2 >= 3). Person-specific classification models were evaluated using an AUC statistic [22]. To assess the robustness of the predictions we used a repeated sampling approach (100 random training-test data splits), where each sample included a 70/30 split of training and test data. For the marginal model (whole sample), train/test splits used subject-wise data splitting [23,24] to avoid overestimating model performance. We also investigated if the algorithm performance improved by learning from early weeks in the "test data;" i.e., participant phone usage pattern for early treatment (*1-4 weeks*) before it began predicting future mood (*5-12 weeks*) for the marginal model including the entire sample.

The basic train/test approach of the analyses is shown in Figure 2.1. All analyses were done using R [25] and made use of the ranger package [26] for random forests models.

Table 2.1. Passive features generated from phone usage data by Ginger.io app.

Passive Feature	Description
Mobility Distance	Approximate distance in miles covered by the user by foot or by bike on a particular day as determined from location data
Mobility Radius	Approximate radius of an imaginary circle encompassing the various locations that a user has traveled across on a particular day, in miles
Call Duration	Total duration of all calls in seconds
SMS Count	Number of SMS messages sent and received
SMS Length	Total length of all SMS messages in□ characters
Aggregate Communication	Total number of calls and total number of□ SMS messages on a particular day
Interaction Diversity	Total number of unique individuals with whom a □participant interacted through phone calls or SMS messages on a particular □day
Missed Interactions	Total number of calls unanswered for a user on a □particular day
Unreturned calls	The number of missed calls without an associated call back

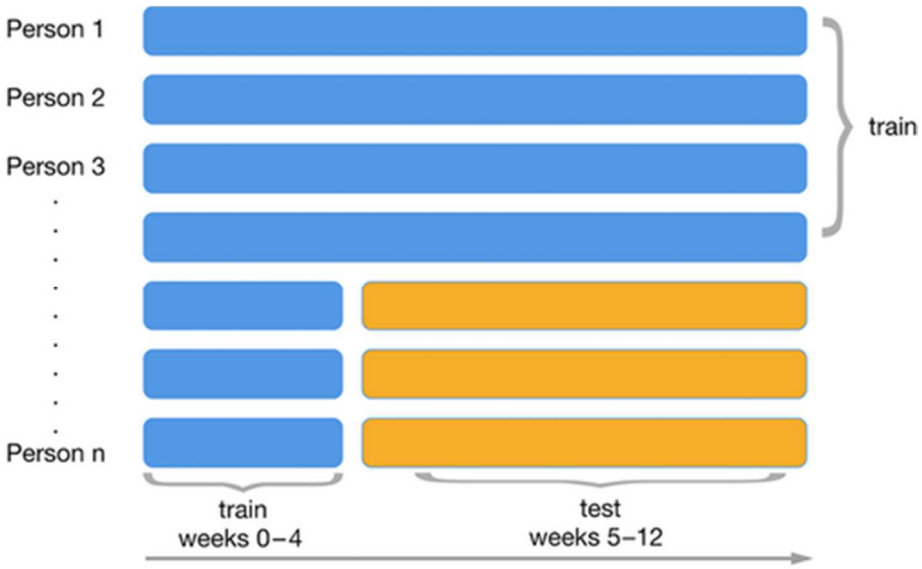


Figure 2.1. Schematic of overall data analysis strategy

For 70% of users, all data was used for training, whereas in the 30% of users in the test set, a variable amount of initial weekly data was also used for training, and the latter data (yellow) was used to test the model predictions.

2.4 RESULTS

2.4.1 *Data summary*

The present sample includes a subset of participants from the original BRIGHTEN sample with Android phones ($N = 271$) that allowed broader array of passive features (calls, messages and GPS) to be compared with PHQ-2. Figure 2.2 shows the summary distribution of select few passive features, and Table 2.2 the summary statistics for all collected passive features.

The average age of the sample was 33.4 years ($SD = 10.7$) and 77.8% of participants were female. The cohort was 57.5% Non-Hispanic White, 16.2% African American/Black, and 15.1% Hispanic. A significant proportion of the participants (35.2%) reported making under $30,000 annually, and a majority (54.2%) said they couldn't make ends meet with their current income. Daily reported

mood using the modified PHQ-2 was 4.48 (*SD* = 2.3, range: 2-10), with wide variability within and between participants. Figure 2.3 shows three different mood trends from six select participants. Participant attrition was linear (Figure 2.4) over the study period. There was no direct association between attrition and assessment incentives at weeks 4, 8, and 12. We considered a participant "active" during the week if any passive or active data was recorded at least once.

Table 2.2. Passive data summary statistics

Statistic	Mean	St. Dev	Min	Pctl(25)	Median	Pctl (75)	Max
Unreturned calls	0.88	1.62	0	0	0	1	27
Missed interactions	1.31	2.36	0	0	1	2	76
Mobility (miles)	1.32	1.34	0.00	0.39	1.00	1.84	18.28
Call count	5.59	7.81	0	1	3	7	97
Interaction diversity	5.99	4.97	0	3	5	8	52
Mobility radius (miles)	14.18	111.83	0.00	0.64	3.59	8.22	7,012.50
SMS count	38.58	66.33	0	4	17	45	1,507
Aggregate communication	44.21	68.06	0	8	23	53	1,510
Call duration (seconds)	1,425.69	2,896.46	0	32	383	1,537	58,334
SMS Length (characters)	1,872.75	3,139.62	0	218	844	2,170	47,741

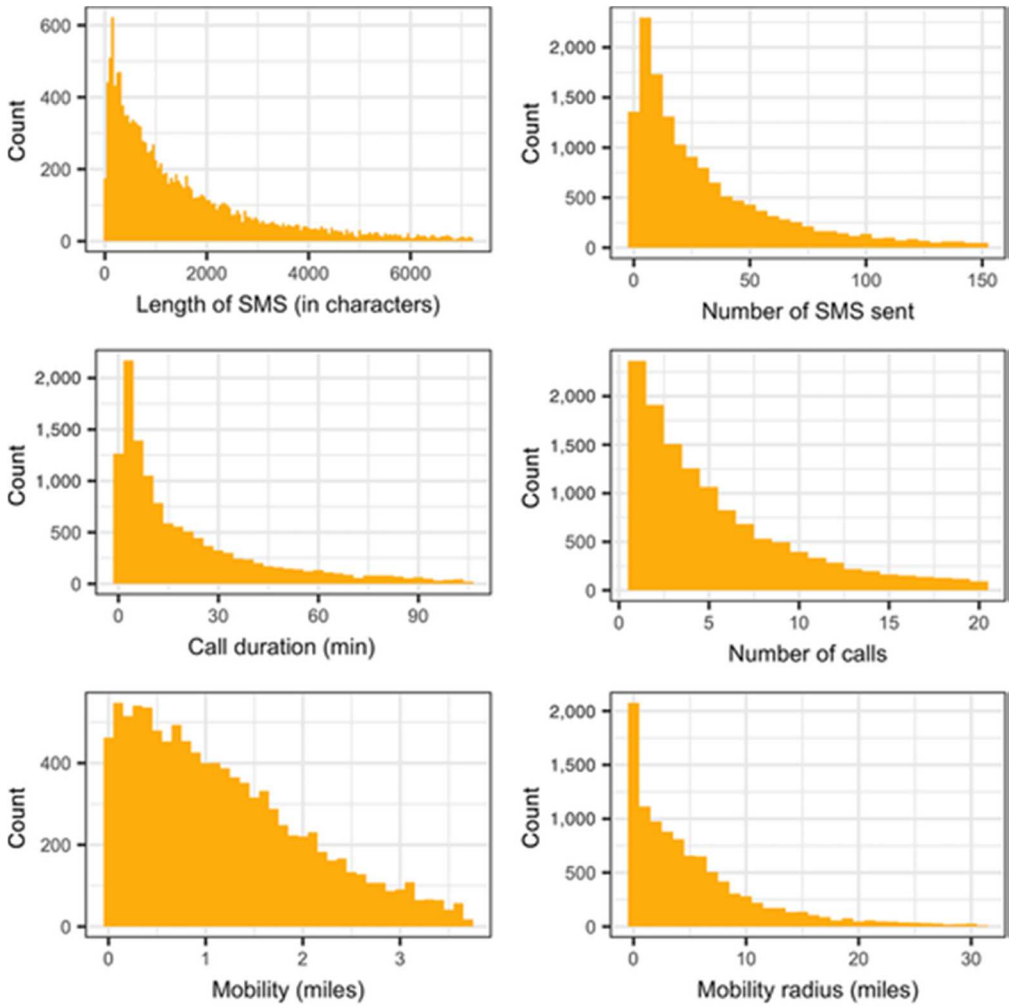


Figure 2.2. Histograms of select passive features as collected from the study cohort. For plotting purposes, the data from lower and upper 5% quantile tails were filtered

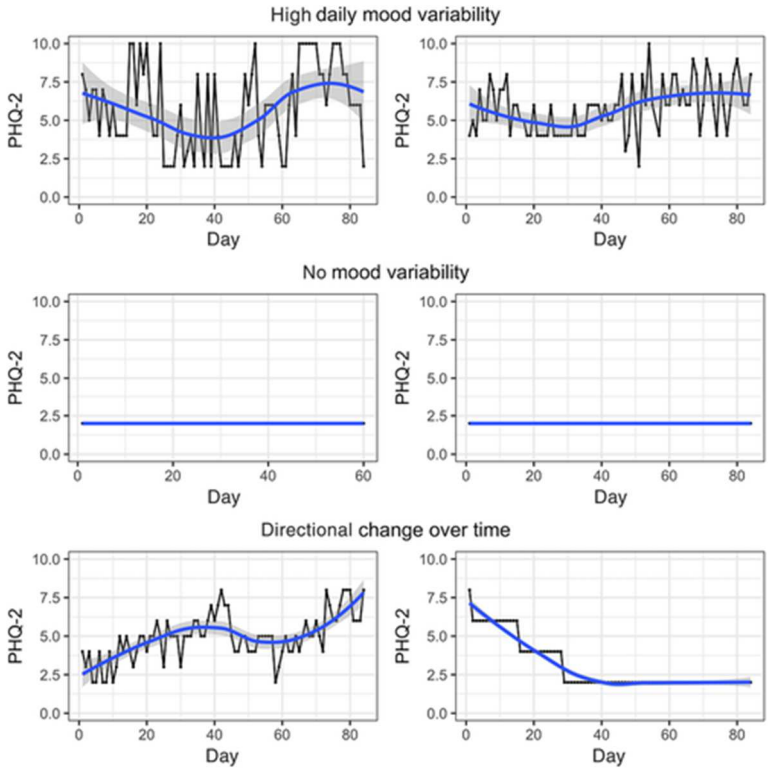


Figure 2.3. Variations in daily self-reported mood of a select few individuals.

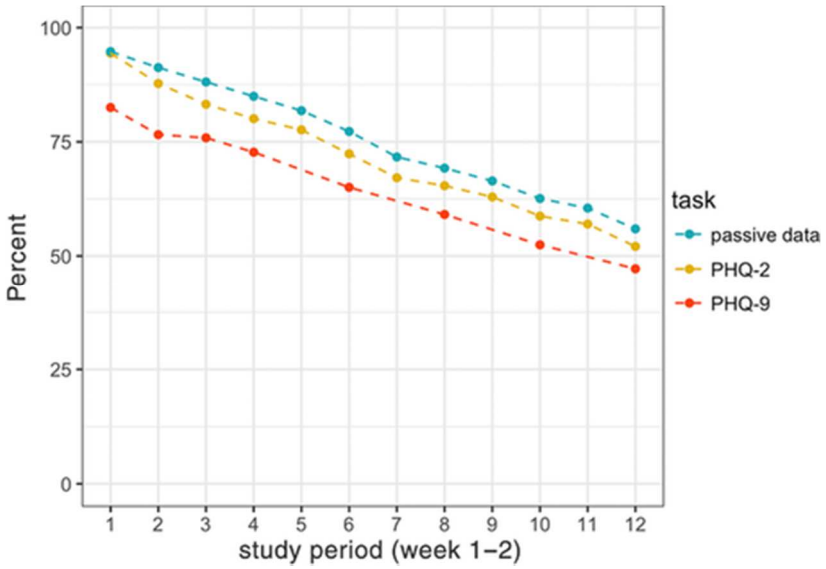


Figure 2.4. Overall participant retention rate in the study.

It is stratified by different kinds of data collected through active tasks (self-reported PHQ-2 and PHQ-9 surveys) and passive phone usage (passive data).

2.4.2 *Association between self-reported daily mood and phone usage*

Pairwise correlations amongst passive features showed three clusters based on mobility, phone usage logs and missed calls. Overall, no significant correlations were found (Figure 2.5) between passive features and PHQ-2. To account for within-subject correlations for longitudinal responses, we used a marginal GEE model with a first order autoregressive working correlation structure. A limited association between PHQ-2 and GPS derived mobility was seen (p = .04). Call count, number of SMS sent, and other derived features showed non-significant borderline association (p < .10) with PHQ-2 (see Table 2.3).

Table 2.3. Model Estimates and standard error of passive data features using a GEE model

	Model estimates (SE)
(Intercept)	4.36 (0.38) ***
Unreturned calls	-0.01 (0.02)
Mobility	-0.04 (0.02) *
SMS length	0.00 (0.00)
Call duration	0.00 (0.00)
Interaction diversity	-0.01 (0.01) .
Missed interactions	0.02 (0.01) .
Aggregate communication	-0.05 (0.03) .
SMS count	-0.06 (0.03) .
Mobility radius	0.00 (0.00)
Call count	0.06 (0.03) .
Age	0.01 (0.01)
GenderMale	-0.06 (0.25)
***p < .001, **p<.01, *p <.05, . p <.1	

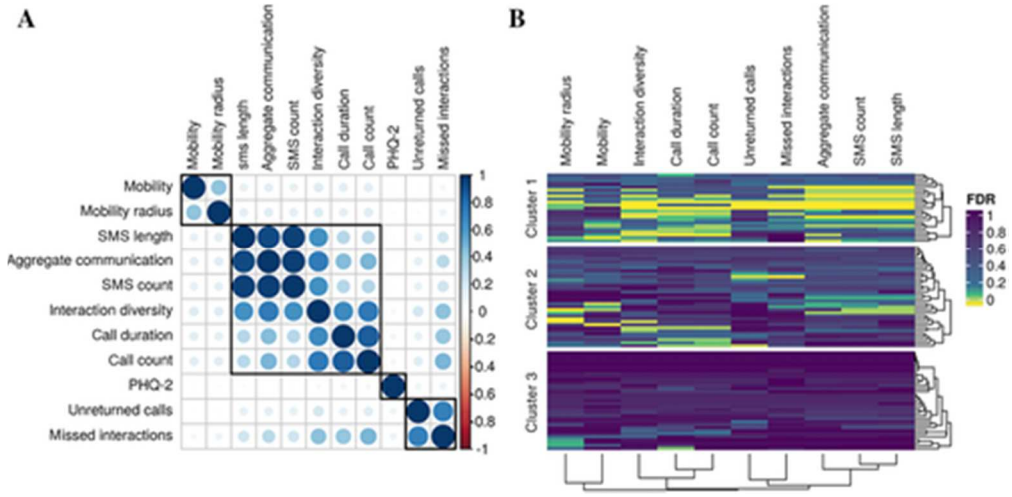


Figure 2.5. Correlation between passive data and association with daily mood at an individual level

(a) A correlation plot of pairwise Spearman correlations between passive features and self□reported mood (PHQ□2) at the cohort level. (b) Personalized (N□of□1) Spearman correlations *P*□values (FDR corrected) between self□reported mood (PHQ□2) and passive features. Cluster 1 shows individuals that have a broad association between the majority of the passive features and daily mood, cluster 2 highlights a subset of individuals that show a weaker, non uniform association between passive features and daily mood, and cluster 3 demarcates a subgroup of individuals that show no relationship between daily mood and passive tracking of phone usage

2.4.3 *Predicting daily mood (PHQ-2) from daily phone usage*

Using the random forest approach, three models were fit utilizing demographics, baseline PHQ-9, and passive phone usage features additively. Prediction results are shown in Figure 2.6 for the three models by number of weeks of additional training data on the test set. Several interesting patterns appear. First, the results at week 0 reflect models developed on 70% of participants, which are then tested on the remaining 30% of participants, without any additional training on this 30%. These models are uniformly poor with median R^2 close to zero. Second, all models get

progressively better (i.e., increasing R^2) with additional weeks of preliminary data from test set data. Note that this pattern is also true for models including only baseline covariates, which may indicate that these models are more accurately learning an individual's stable (i.e., mean) mood with additional weeks of training data. Taken together, these results suggest that whatever associations there are between predictors and mood, they tend to be fairly unique to individuals. Finally, contrary to hypothesis, the passive phone features do not enhance prediction, over and above demographics and baseline PHQ-9. For these marginal results, the passive phone features appear to worsen prediction.

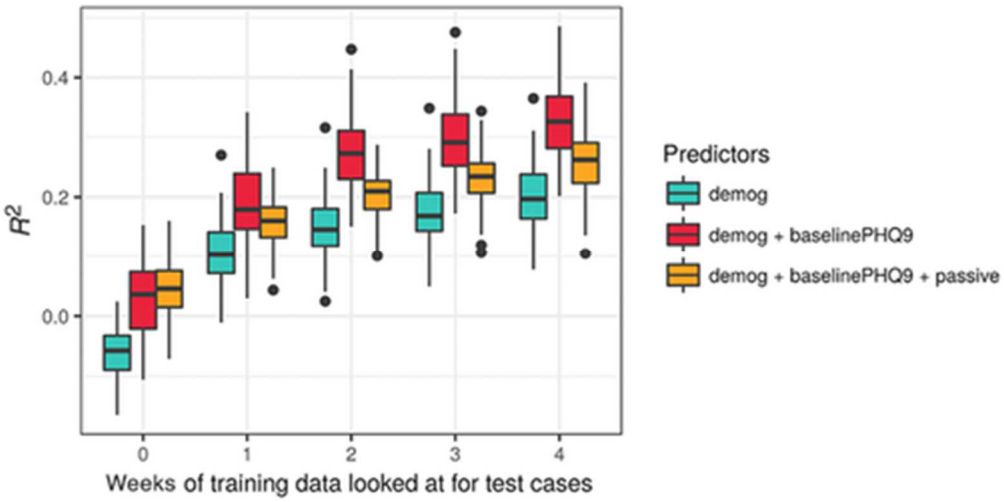


Figure 2.6. Comparison of random forest prediction models based on R^2 for different feature sets. The x-axis shows the performance of iterative model retraining using the data from test users for week 0 (no test data used for training) to 1–4

2.4.4 *Personalized mood prediction*

A subset of 93 participants were selected for individual prediction models based on the following criteria: a) variability in daily mood (an interquartile difference in PHQ-2 of at least 1), b) distribution of class labels (minimum of 20% in mild or severe state), and c) at least 15 days of

longitudinal data. These individual-level correlations showed significant heterogeneity between passive features and daily mood (Figure 2.5b). A subset of 13.9% of these participants showed significant association (FDR < .10) between one of the passive features and daily mood. The random forest based classification was able to predict PHQ-2 state better than chance for 80.6% of individuals (75 out of 93; median AUC > .50, 100 random splits) from passive features alone. Eleven individuals had median AUC greater than .80, demonstrating high predictive power in inferring daily mood from phone usage patterns. To assess the sensitivity of our predictions we shuffled true PHQ-2 state labels. The overall trend between true and shuffled response (Figures 7b and 7c, respectively) shows a viable signal in the passive data for predicting PHQ-2. However, power is greatly reduced by running individual prediction models, as seen in permutation-based tests. Ensemble methods like random forests do not lend themselves to straightforward interpretation of predictors (e.g., there are no regression coefficients), but it is possible to examine which predictors appear most "important" in the prediction, using the Gini index [27]. While no passive feature uniformly stood out, GPS-based mobility distance and mobility radius were the top two predictors of daily PHQ-2. The heatmap display of predictor importance (Figure 2.7. d) highlights the heterogeneity of passive features for predicting PHQ-2 across individuals. For illustrative purposes, Figure 2.8 shows daily PHQ-2 and passive data for a select individual with >0.9 median AUC score.

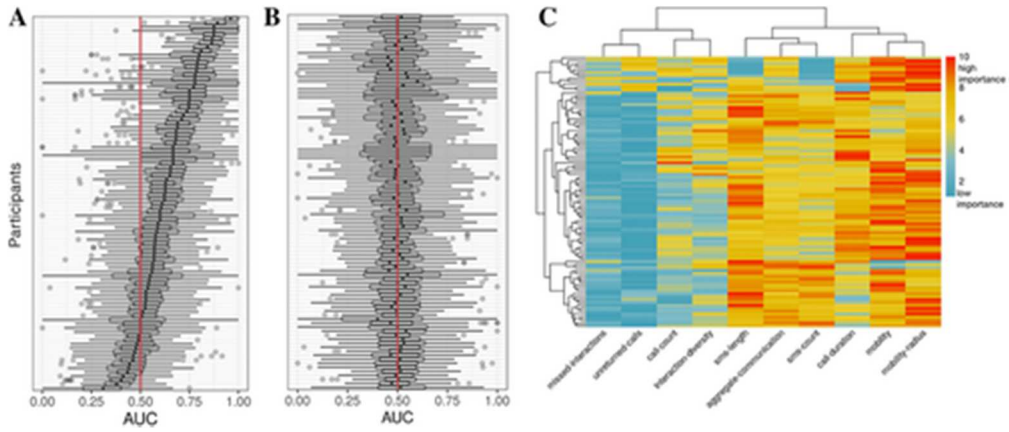


Figure 2.7. Performance of personalized models evaluating the ability to predict daily mood (a) Distribution of area under the curve (AUC) scores for an individual's daily mood prediction (low/high) based on random forest models (N = 93). (b) Null distribution for AUC scores based on shuffled daily mood labels. Red line indicates AUC = 0.50 (equivalent to a probability of random coin toss). (c) Top predictive features based on average ranks (1 = low importance and 10 = high importance) of variable importance derived using Gini index impurity scores from personalized random forest models

2.5 Discussion

Our findings are particularly relevant given the upsurge of interest in using digital technologies to augment the clinical care of depression. Specifically, our study shows that passive data offers the most promise in predicting depressive symptoms at an individual level, whereas there is little evidence for an overarching prediction algorithm that is applicable to a wide variety of individuals While some smaller studies [10,28,29,30] have found associations between passive mobility data and severity of depressive symptoms, our examination of a large, nationally recruited sample of individuals with depressive symptoms did not show a meaningful relationship between phone usage features and daily mood at the cohort level. We believe further large-scale studies (N > 10,000) and longer data collection (> 12 weeks) are needed to stratify robust signatures of digital phenotypes from passive data and contextualize their association to mood. Our findings indicate

GPS mobility may have the greatest potential to harness mobile technology to infer mood. Previous demonstrations ([10,12] used GPS data from smartphones to predict depressive symptoms based on features such as location preference and mobility patterns. However, these demonstrations were on smaller samples (< 40 individuals) and represent only the first steps in understanding the ability of passive mobility data to make inferences about depressive symptoms.

In the context of these mixed findings, our results shed light on the potential for smartphones in measurement-based care for depression. Notably, our data highlight that mood states are best predicted at an individualized level by looking at one's own deviation than by comparing one against a population norm. We also observed a high degree of intra- and inter-individual variance in daily phone usage and mood ratings. This reinforces the notion that optimal clinical decision making for depression should be based on more regular monitoring of symptoms and treatment outcomes, rather than infrequent self-reports obtained at clinic visits. Measurement-based care is intended to provide feedback to both patients and providers about treatment response and navigate treatment goals accordingly; when done well, such measurement may facilitate patient-provider communication and shared decision making [31]. Future trials may consider the relative importance of various types of passively collected data for depression care; for example, the role of both overall mobility and specific location (e.g., home, work, recreation) for behavioral activation and the importance of phone usage to monitor social engagement. Moreover, measurement-based care is best integrated with existing clinic infrastructures (e.g., electronic health records) to alleviate the burden of routine collection and supplement clinical decision making. Although smartphone technology may allow for novel methods of data collection, future research is needed to better integrate such passive data collection into existing clinic structures and processes [32].

Despite the promise, the clinical utility of passive sensing to predict a person's mood and overall behavioral health has a long way to go (Renn et al., 2018). There are several constraints that mHealth researchers should be aware of: 1) *Learning robust, generalizable behavior patterns from data*: With intensive longitudinal data from phone usage, a risk is that we learn highly idiographic associations between passive phone features and daily mood, whereas the overarching research goal is to learn about generalizable patterns that are applicable to populations of individuals. Care is needed in running and evaluating machine learning models to avoid learning idiosyncratic digital fingerprints rather than broadly applicable associations of the passive features with mood fluctuations [33]. 2) *Platform heterogeneity*: Significant differences between iOS and Android platforms impact passive data features, granularity, and sampling rate. iOS, for example, restricts acquisition of phone and messaging logs. 3) *Passive data*: Until we are able to reliably infer behavior patterns from passive data features, raw data sampled at high frequency should be stored and analyzed, rather than proprietary summary statistics from apps. Further work is needed to explore new passive features such as number of notifications accepted, screen usage, mobile apps used in a day, total keyboard strokes, reaction time, etc. [34]. The interplay of these features may help build robust digital phenotypes of mood. 4) *Data contextuality*: Context, quality and quantity of user interaction with smartphones may be meaningful. [13] recently showed the importance of the nature of an individual's location (e.g., house of worship, recreation) to better understand how the contexts of physical locations relate to depression. Similarly, missed and unreturned calls from friends and family should be weighted more heavily in comparison to missed calls from unknown numbers. 5) *User engagement* - To enable robust learning from the rich but noisy continuous passive data streams, requires both deep (e.g., number of days) and large (e.g., number of users)

data. Although large mHealth [35] studies can provide powerful means to recruit a large number of individuals, there are significant challenges in user retention and compliance with study protocols that often result in sparse data collection. We believe mHealth apps that empathize with users, address their daily needs and clearly articulate data security, sharing and research usage policy will help gain long-term user trust and engagement.

Strengths of the present study leverage those of the BRIGHTEN parent study. We recruited one of the largest samples to date investigating passive phone data and mood. Furthermore, the present analysis applied machine learning to learn nonlinear behavior patterns from phone data to predict daily mood. Nonetheless, our findings must be considered in light of limitations. Smartphone-specific operating system limitations restricted our analyses to Android users only. We found that data acquisition was easier than the analysis and, in some cases, analysis was prohibited due to data sparsity. The average participant retention rate (51.74% after 12 weeks of study completion) was significantly higher than other recent mobile health studies (7-15%) [35–37]. However, the present study offered financial incentives to participants for each completed assessment, which is not typical of these other studies and likely influenced engagement and retention in the study. Finally, the present sample is not necessarily representative of the underlying population of adults with mild-to-moderate depressive symptoms. Notably, our study was majority female, although this corresponds to the greater prevalence of depressive disorders in women relative to men [38].

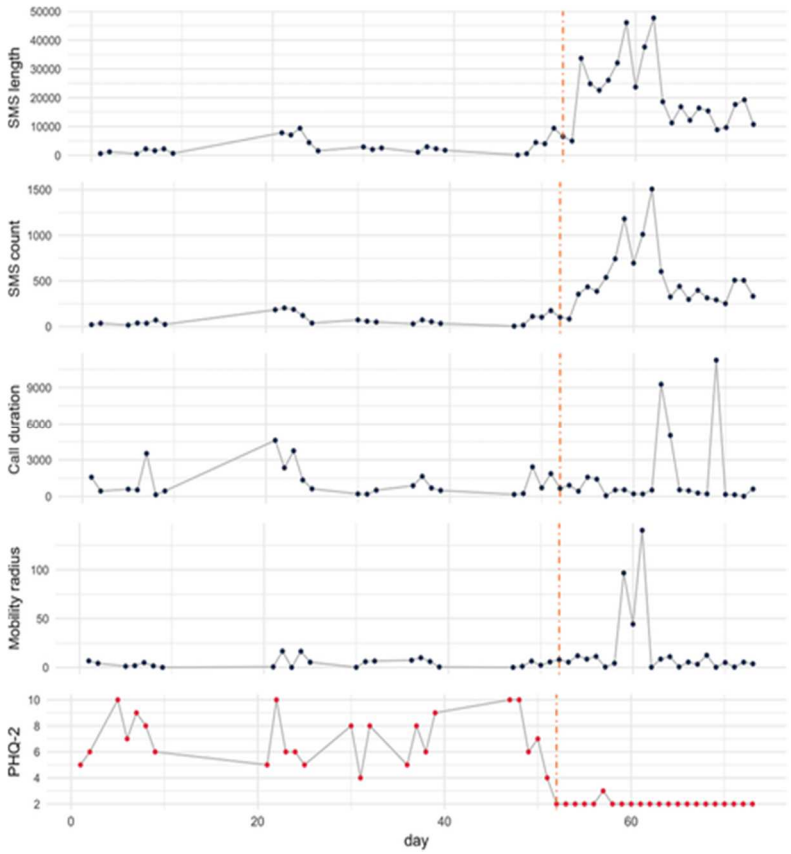


Figure 2.8. Comparison of daily mood and a passive features of an individual participant in the study

2.6 CONCLUSION

Passive data streams from phones offer a potentially unobtrusive way to facilitate clinical care for depression by assessing treatment response and triggering follow-up assessment and treatment modification. Using readily available smartphone technology facilitates the scalability of such approaches at a fraction of the cost of in-clinic visits. There is growing preliminary evidence that daily mobility patterns obtained from phone sensors are associated with depressive symptom severity, although this is most salient when assessing individual change over a course of treatment. Additional large-scale studies ($N > 10{,}000$) with long-term user engagement are needed to uncover

the passive features best suited to detecting and monitoring changes in depressive symptoms and related functioning.

2.7 REFERENCES

1. Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).

2. National Institute of Mental Health, 2018. Available at: https://www.nimh.nih.gov/health/topics/depression/index.shtml. (Accessed: 30th March 2017)

3. [No title]. Available at: http://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf. (Accessed: 26th November 2017)

4. Areàn, P. A., Hoa Ly, K. & Andersson, G. Mobile technology for mental health assessment. *Dialogues Clin. Neurosci.* **18**, 163–169 (2016).

5. Simon, G. E. *et al.* Does Response on the PHQ-9 Depression Questionnaire Predict Subsequent Suicide Attempt or Suicide Death? *Psychiatr. Serv.* **64**, 1195–1202 (2013).

6. Passini, C. M., Pihet, S., Favez, N. & Schoebi, D. Ecological Momentary Assessment Parenting Scale. *PsycTESTS Dataset* (2013). doi:10.1037/t38554-000

7. Torous, J., Kiang, M. V., Lorme, J. & Onnela, J.-P. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Ment Health* **3**, e16 (2016).

8. Onnela, J.-P. & Rauch, S. L. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology* **41**, 1691–1696 (2016).

9. Haftor & Darek. *Information and Communication Technologies, Society and Human Beings: Theory and Framework (Festschrift in honor of Gunilla Bradley): Theory and Framework (Festschrift in honor of Gunilla Bradley)*. (IGI Global, 2010).

10. Saeb, S. *et al.* Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J. Med. Internet Res.* **17**, e175 (2015).

11. Burns, M. N. *et al.* Harnessing context sensing to develop a mobile intervention for depression. *J. Med. Internet Res.* **13**, e55 (2011).

12. Canzian, L. & Musolesi, M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15* 1293–1304 (ACM Press, 2015).

13. Saeb, S., Lattie, E. G., Kording, K. P. & Mohr, D. C. Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety. *JMIR Mhealth Uhealth* **5**, e112 (2017).

14. Anguera, J. A., Jordan, J. T., Castaneda, D., Gazzaley, A. & Areán, P. A. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. *BMJ Innov* **2**, 14–21 (2016).

15. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).

16. Löwe, B., Kroenke, K. & Gräfe, K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J. Psychosom. Res.* **58**, 163–171 (2005).

17. Liang, K.-Y. & Zeger, S. L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **73**, 13 (1986).

18. Ballinger, G. A. Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods* **7**, 127–150 (2004).

19. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

20. Chen, X. & Ishwaran, H. Random forests for genomic data analysis. *Genomics* **99**, 323–329 (2012).

21. Boulesteix, A.-L., Janitza, S., Kruppa, J. & König, I. R. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2**, 493–507 (2012).

22. Beck, J. R. & Shultz, E. K. The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch. Pathol. Lab. Med.* **110**, 13–20 (1986).

23. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C. & Kording, K. P. The need to approximate the use-case in clinical machine learning. *Gigascience* **6**, 1–9 (2017).

24. Elias Chaibub Neto, Abhishek Pratap, Thanneer M Perumal, Meghasyam Tummalacherla, Brian M Bot, Lara Mangravite, Larsson Omberg. Detecting confounding due to subject identification in clinical machine learning diagnostic applications: a permutation test approach. *arXiv* (2017).

25. Website. Available at: R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (Accessed: 31st August 2017)

26. Wright, M. N. & Ziegler, A. ranger: A Fast Implementation of Random Forests for High

Dimensional Data in C and R. *J. Stat. Softw.* **77**, (2017).

27. Strobl, C., Boulesteix, A.-L. & Augustin, T. Unbiased split selection for classification trees based on the Gini Index. *Comput. Stat. Data Anal.* **52**, 483–501 (2007).

28. Ghandeharioun, A. *et al.* Objective Assessment of Depressive Symptoms with Machine Learning and Wearable Sensors Data. *ACII2017* (2017).

29. Wang, R. *et al.* StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct* 3–14 (ACM Press, 2014).

30. Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M. & Weidt, S. Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. *JMIR Mhealth Uhealth* **4**, e111 (2016).

31. Scott, K. & C.Lewis, C. Using Measurement-Based Care to Enhance Any Treatment. *Cogn. Behav. Pract.* **22**, 49–59 (2015).

32. Hallgren, K. A., Bauer, A. M. & Atkins, D. C. Digital technology and clinical decision making in depression treatment: Current findings and future opportunities. *Depress. Anxiety* **34**, 494–501 (2017).

33. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C. & Kording, K. P. The need to approximate the use-case in clinical machine learning. *Gigascience* **6**, 1–9 (2017).

34. Mehrotra, A. *et al.* Understanding the Role of Places and Activities on Mobile Phone Interaction and Usage Patterns. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **1**, 1–22 (2017).

35. Dorsey, E. R. *et al.* The Use of Smartphones for Health Research. *Acad. Med.* **92**, 157–160 (2017).

36. Chan, Y.-F. Y. *et al.* The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. *Nat. Biotechnol.* **35**, 354–362 (2017).

37. McConnell, M. V. *et al.* Feasibility of Obtaining Measures of Lifestyle From a Smartphone App: The MyHeart Counts Cardiovascular Health Study. *JAMA Cardiol* **2**, 67–76 (2017).

38. Kessler, R. C. *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* **289**, 3095–3105 (2003).

Chapter 3. FEASIBILITY OF UTILIZING TECHNOLOGY TO ASSESS AND INTERVENE IN DEPRESSION IN HISPANICS AND LATINOS

3.1 ABSTRACT

Background:

Most people with mental health disorders fail to receive timely access to adequate care. In the U.S., Hispanic/Latino individuals are particularly underrepresented in mental health care and are historically a very difficult population to recruit into clinical trials. However, U.S. Hispanic/Latinos have increasing access to mobile technology, with over 75% owning a smartphone. This technology has the potential to overcome known barriers to accessing and utilizing traditional assessment and treatment approaches.

Objective:

This is a feasibility clinical trial comparing three different types of mental health apps for the treatment of mild to moderate depression in Hispanic/Latino adults in the U.S. The primary aim of this study was to to compare recruitment and engagement in a fully remote trial of individuals with depression who either self-identify as Hispanic/Latino or not. A secondary aim was to assess treatment outcomes from three different self-guided mobile apps (iPST (based on evidence-based therapeutic principles from problem-solving therapy [PST])); Project: Evolution™ (EVO; a cognitive training app based on cognitive neuroscience principles); and Health Tips (health information app that served as an information control).

Methods:

Spanish- and English-speaking participants were recruited through social media platforms, internet-based advertisements, and traditional fliers in select locations in each state across the United States. Assessment and self-guided treatment was conducted on each participant's smartphone or tablet. We enrolled 389 Hispanic/Latino and 637 non-Hispanic/Latino adults (≥18 years old) with mild to moderate depression as determined by a 9-item Patient Health Questionnaire (PHQ-9) score ≥5 or an endorsement of impaired functioning. Participants were first asked their preferences among the three apps available to them, and then randomized to their top two choices. Outcomes were depressive symptom severity (measured using PHQ-9) and functional impairment (assessed with Sheehan Disability Scale) and collected over 3 months. Engagement in the study was assessed based on the number of times participants completed active surveys.

Results:

We screened 4,502 participants and enrolled 1,040 participants from throughout the U.S. over 6 months, yielding a sample of 348 active users. The majority of the participants were recruited via posts on craigslist.org, with significant acquisition costs for recruiting Spanish-speaking Hispanic/Latinos participants ($31/ participant) compared to their English-speaking non-Hispanic/Latino counterparts ($1.49/participant). Long-term engagement surfaced as a key issue among Hispanic/Latino participants, who dropped from the study two weeks earlier than their non-Hispanic/Latino counterparts ($p < 0.016$). There were no significant differences observed for treatment outcomes between those identifying as Hispanic/Latino or not. Although depressive symptoms improved over the course of treatment, outcomes did not vary by type of treatment app.

Conclusions:

The findings from this study suggest that fully remote mobile-based studies can attract a diverse participant pool including people from traditionally underserved communities in mental health care and research (in this case, Hispanic/Latino individuals). However, keeping participants engaged in this type of 'low-touch' research study remains challenging. Hispanic/Latino populations may be less willing to use mobile apps for assessing and managing depression. We recommend that future research endeavors include the use of user-centered design to determine the role of mobile apps in assessment and treatment of depression for this population, app features they would be interested in using, and strategies for long-term engagement.

Trial Registration: Clinicaltrials.gov Identifier: NCT01808976

3.2 INTRODUCTION

Technology is being leveraged as a way to do large-scale clinical research targeting typically underrepresented populations. Given the extensive use of mobile devices across communities, remote research methods are becoming widely used. Additionally, technology is also seen as a potential method for bridging health disparities, which are typically driven by limited resources and stigma most apparent in minority communities. Of particular interest is the Hispanic/Latino community: Although they comprise one of the fastest-growing demographic segments in the U.S.[1], Hispanic/Latino populations in the U.S. are half as likely as their non-Hispanic White counterparts to receive mental health services[2]. This population is very difficult to recruit into research[3,4] and as a result, there is limited science to support treatment recommendations for this population. Recruitment of Hispanic/Latino samples into clinical research is particularly challenging in studies of mental health.

The widespread availability of digital technology has the potential to drive a sea change in access to psychosocial treatment for mental health problems in Hispanic/Latino communities[5]. Internet-based interventions have already demonstrated comparable treatment outcomes as traditional face-to-face psychotherapy[6], and given that 75% of Hispanic/Latino own a smartphone[7], mobile-based mental health applications ("apps") have the potential to increase treatment accessibility and engagement. Although there is potential for treating depression in Hispanic/Latino individuals using mobile devices, there is relatively little information about how this population interacts with apps, given their underrepresentation in mental health research. In particular, do Hispanic/Latino smartphone owners (including both Spanish- and English-speakers) actually use mental health apps, and when they do, do they follow the app protocols? We recently tested similar questions among a majority non-Hispanic White sample in a recent, fully remote trial (BRIGHTEN V1; [8,9]) and found that interest in depression apps was high. It was far less challenging to recruit participants into our remote clinical trial compared to traditional in-person treatment trials. However, long-term engagement with the assigned apps trailed off significantly each week in the study; a finding that has been demonstrated in other studies[10]. However, Hispanic/Latino individuals, especially non-English speakers, do not typically have the same opportunity as majority groups to utilize mental health services and therefore may find mental health apps a useful alternative to traditional care. There is an immediate need for further research to develop and evaluate new solutions for mental health care for this population that are economically viable, scalable, and focused on engaging users to inform timely and evidence-based clinical interventions.

Therefore, the aim of this study was to determine the feasibility of conducting remote research with a Hispanic/Latino adult sample of smartphone users, how they interact with depression apps,

and the potential clinical impact mobile health apps may have on treating depression in this population. We report recruitment, engagement, and cost in this 12-week fully remote randomized controlled trial among Hispanic/Latino individuals with depression and a cohort of non-Hispanic/Latinos with depression to act as a direct comparator group (and extend our previous findings).

3.3 METHODS

Ethical approval for the trial was granted by the UCSF Committee for Human Research. Specific research methods for this project replicated the BRIGHTEN V1 study and are described elsewhere[8,9], but are summarized here. Briefly, this was fully remote treatment trial for depression, consisting of engagement with one of three treatment apps and periodic assessments detailed below.

3.3.1 *Recruitment*

Three different types of recruitment approaches, including traditional, social networking, and search-engine strategies, were used (Figure 3.1). Traditional methods consisted of craigslist.org postings throughout the United States, specifically posting to the 'Volunteer' and 'Jobs etc.' pages within Craigslist in at least one major city in every state. Social networking methods included regular postings on sites such as Facebook and Twitter, and contextual-targeting methods to identify and directly push recruitment ads to potential participants, based on their Twitter and other social media comments. This approach was led entirely by trialspark.com, which designed specific recruiting campaigns using machine learning approaches to create optimal advertising. Furthermore, we reached out to Hispanic/Latino Catholic Ministries in at least one city in every

state to see if they would be willing to champion this study and post flyers in their communities. Each approach provided potentially interested participants a link to our custom study website (www.brightenstudy.org), which was translated entirely for Spanish-speakers (www.brightenstudy.org/spa) and included a welcome video featuring bilingual Hispanic/Latino researchers describing the goal of this study (http://bit.ly/2D06KKK) in Spanish. All translations involving text in the treatment apps were done by a combination of native Spanish speakers associated with this study and professionals at Babble-on (https://www.ibabbleon.com/translation.html).

3.3.2 *Procedures*

This study used an equipoise stratified clinical trial design [11], which factors participant preferences for treatment into the randomization. Participants were randomly assigned one app amongst their two preferred intervention types and asked to use it daily for 4 weeks. Participants completed primary outcome assessments (PHQ-9[12], Sheehan Disability Scale (SDS)[13]) once a week for 3 months, with other secondary measures (described below) completed at daily, weekly, or biweekly intervals. All treatment and assessment was delivered remotely via custom apps.

Screening: Interested participants completed a brief online screening consisting of questions about their ability to speak Spanish ("Do you speak Spanish? (¿Hablas Español?)") and mobile device ownership ("Do you have an iPhone or Android smartphone?").

*Consent***:** Participants were given a PDF of the UCSF consent form to read, and were instructed to watch a video that highlighted the goals and procedures of the study, as well as risks and benefits of participation. After viewing the video, participants had to pass a quiz that confirmed their understanding that participation was voluntary, was not a substitute for treatment, and that they

were to be randomized to treatment conditions. Each question had to be answered correctly before moving on to baseline assessment and randomization. Eligibility was established after consent was obtained. Upon being eligible, participants were sent a link to download their assessment application (Surveytory).

3.3.3 *Participant eligibility*

Participants had to speak English or Spanish, be 18 years old or older, and own either an a) iPhone with Wi-Fi or 3G/4G/LTE capabilities or b) an Android phone along with an Apple iPad version 2.0 or newer device. iOS based device was required as one of our intervention apps were only available on iOS devices at the time of the study. If a user had an Android phone, they were only eligible to participate if they also owned an Apple iPad version 2 or newer iOS tablet device. Participants had to endorse clinically significant symptoms of depression, as indicated by either a score of 5 or higher on the Patient Health Questionnaire ([PHQ-9]), or a score of 2 or greater on PHQ item 10 (indicating that they felt disabled in their life because of their mood).

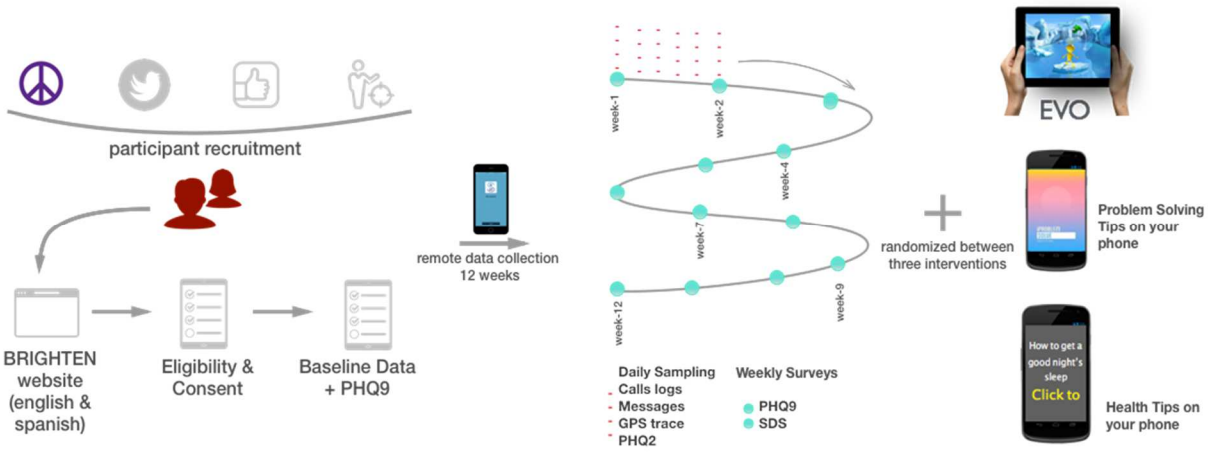


Figure 3.1. Overall Brighten V2 study schematic

The flow shows participant recruitment, consent, enrollment and randomization workflow along with weekly and daily data collection.

3.3.4 *Assessment*

Baseline: The baseline assessment included the collection of demographic variables including age, race/ethnicity, marital and employment status, income, education, smart device ownership, use of other health apps, and use of mental health services, including use of medications and psychotherapy. We collected information on mental health status using the PHQ-9[12] for depression and the Sheehan Disability Scale (SDS)[13] to assess self-reported disability. The PHQ-9 rates the presence and severity of depressive symptoms across nine items, with higher scores signifying more severe symptomatology (range 0-27). This is a reliable and well-validated screening instrument[14] that is responsive to depression treatment outcomes over time[12], and is included in the U.S. Preventive Services Task Force recommendation for depression screening in adults [15]. The PHQ-9 has been translated into several languages; we used both the original English language form and the validated Spanish translation[16]. The baseline PHQ-9 demonstrated good internal consistency in our sample (Cronbach's alpha = 0.85, 95% confidence interval 0.83-0.87). The SDS

assesses perceived functional impairment across three domains (work/school, social life, and family/home responsibilities), yielding a sum score (0-30) in which higher scores represent greater disability. The SDS is popular in clinical trials given its sensitivity in detecting treatment effects[17]. As one of the official World Health Organization's measure of disability, this measure has also been translated into several languages; we used both the original English version and a validated Spanish translation of this scale [18]. The SDS also demonstrated good internal consistency in our sample (Cronbach's alpha = 0.89, 95% confidence interval 0.87-0.91).

Follow up assessments: Our custom mobile app, Surveytory, was used to collect all outcome and passive data. The assessments to measure changes in mood (PHQ-9) and disability (SDS) were administered weekly. Daily changes in mood were assessed using a PHQ-2 survey. Passive data collection included daily phone usage logs (call/text time, call duration, and text length) and mobility data (activity type and distance traveled using the phone's accelerometer and GPS). Participants were automatically notified every 8 hours for 24 hours if they had not completed a survey within 8 hours of its original delivery. A built-in reminder also prompted the participant to check for any surveys on a daily basis in case they missed a new survey notification. An assessment was considered missing if it was not completed within a 24-hour time frame.

Treatment: After confirming completion of baseline assessments (or 72 hours after the initiation of these assessments, whichever came first), participants were sent an online survey which described each of the three treatment arms. Following this description, participants were asked to select which two apps they were most inclined to use in this study. Participants were then randomly assigned to one of these two preferred conditions and sent a link to download the intervention app, which included a brief video explaining how to download and use the assigned treatment app. This download also included a custom dashboard to monitor their study progress. Participants were

asked to use their assigned app for one month. The first app was a video game-inspired cognitive intervention (Project: Evolution™ [EVO]) designed to modulate cognitive control abilities, as declines in these abilities have been associated with depression[19]. This intervention has preliminary evidence for being an effective treatment for depression [19]. The second intervention was an app based on problem-solving therapy (iPST), an evidence-based treatment for depression, which has been shown to be both acceptable and efficacious for U.S.-dwelling Hispanic/Latino populations. The final intervention app, an information control, provided daily health tips (HTips) for overcoming depressed mood such as self-care (e.g., taking a shower) or physical activity (e.g., taking a walk; see[9] for further descriptions of each). Each of the three apps represented the most common type of self-guided depression apps available at the time of the study: apps based on psychotherapy principles, apps that claim to improve mood through therapeutic games, and apps that provide suggestions for mindfulness and behavioral exercises. Similar to the assessment notifications, each intervention app was equipped with built-in reminders asking the participant to use their app on a daily basis (reminders were sent once daily).

***Incentives*:** Randomized participants were paid a total of $75 in Amazon gift vouchers for completing all assessments over the 12 weeks. Participants received $15 for completing the initial baseline assessment and an additional $20 for each subsequent assessment at the 4-, 8-, and 12-week timepoints.

Procedures to reduce gaming: "Gaming" is a situation where a user enrolls in a study solely to acquire research payment, or attempts to influence specific methodological aspects of the study. We utilized the following safeguards to prevent this: i) locking the eligibility or treatment randomization survey if a participant tried to change a submitted answer so that only the initial

answer was utilized, ii) using study links that are valid for one user/device, and iii) tracking IP addresses to minimize duplicate enrollment.

3.3.5 *Statistical Analyses*

Participant self-reported race/ethnicity was used to create two groups of Hispanic/Latino and non-Hispanic/Latino (e.g., all other races and ethnicities) to test our main study aims. Sample demographics and clinical characteristics were calculated using appropriate descriptive statistics. Comparison between participant demographics were done using a chi-square test of independence for categorical variables and one-way analysis of variance (ANOVA) to compare continuous variables across the groups. To assess the marginal effect (i.e., association in the entire sample) between longitudinal weekly PHQ-9 and SDS scores and treatment arms, we used generalized estimating equations (GEEs)[20]. Briefly, GEE models extend generalized linear models to longitudinal or clustered data. GEEs use a working correlation structure that accounts for within-subject correlations of participant responses, thereby estimating robust and unbiased standard errors compared to ordinary least squares regression [20,21]. We adjusted for age and gender to account for any potential confounding effects between outcome and main covariates of interest. Treatment response was further categorized into three groups based on a change of at least 5 points on the PHQ-9[12] (the minimal clinically important difference[12]), to comprise treatment responders (decrease PHQ-9 >= 5 points), non-responders (change in PHQ-9 less than 5 points), and those that deteriorated over treatment (increase in PHQ-9 of >= 5 points). To assess participant engagement, we examined the proportion of participants that completed at least one activity in any given week. ANOVA was used to compare the daily, weekly and overall participation differences between Hispanic/Latino and other participants. Univariate estimation of time to drop out from the

study between Hispanic/Latino and non-Hispanic/Latino participants was computed using survival analysis. The distribution of the 'survival' days (total days active in the study) and nonparametric estimates of the survivor function was computed using the Kaplan-Meier method[22], and the log-rank test[23] was used to test for differences in survival between Hispanic/Latino and other participants. To compare drop-out rates among the three interventions, a non-parametric Kruskal-Wallis test was used. Passive data was only used to compare user engagement with active survey-based tasks. Given this study design is similar to that of our previous work [8], we used the same power analysis for this study. It indicated that 200 participants per intervention arm would provide 0.80 power to detect a medium treatment effect (e.g., 2 points change on PHQ-9 scale, Cohen's *d* ~ 0.4) with an assumption of 50% participant dropout. However, this study was a feasibility trial of an understudied Hispanic/Latino population and was not sufficiently powered to detect a moderate effect size across the three interventions. All analyses were carried out using R, statistical computing language version 3.4.2[24].

3.4 RESULTS

3.4.1 *Recruitment and Enrollment*

The BRIGHTEN V2 study started recruitment in August 2016 with screening and enrollment continuing for seven months. A total of 4,502 people were screened and 1040 (23.10%) adults met the eligibility criteria and enrolled in the study. Of these, 389 (37.40%) reported being Hispanic/Latinos. As in BRIGHTEN V1 study[8,9], the use of craigslist.org was the most effective approach in recruiting, with more than 80% of our participants coming from this approach. An additional 8% were referred by friends or colleagues.

Enrolled participants lived throughout the US, with all the metropolitan areas represented (Figure 3.2). Only 348 (33.46%) of the initially enrolled participants were active in the study (active cohort), as defined by completing at least one post-enrollment weekly PHO-9 assessments and/or providing passive phone usage data within the first 12 weeks. The remaining 692 (66.54%) participants did not respond to any post-enrollment surveys or provided passive data and as a result, were considered to be study dropouts (Figure 3.3). Income, education, and race were significantly different between those who dropped and those who did not ($p < .005$). A large proportion of individuals who reported that they "can't make ends meet" with regards to their income, dropped out of the study (34.4%); this effect was more pronounced for Hispanic/Latino individuals (47.7%). Over half (60.4%) of the Hispanic/Latino participants who dropped from the study reported making $20,000 or less annually, compared to 28.1.0% of non-Hispanic/Latinos who dropped. Of the 348 active individuals, 74 did not complete the treatment randomization survey, and thus were not assigned an intervention. However, they continued to complete self-report surveys during the study period. For this reason, we categorized these participants as not randomized (EnR) category. All further analyses were restricted to active individuals consisting of those in treatment (n = 274) or EnR arms (n = 74; total N = 348). See Figure 3.3 for the CONSORT diagram illustrating participant flow through the study.

Of those who were randomized, 31.8% attempted to change their assigned intervention by hitting the 'back' button to return to the randomization page, while an additional 10.4% participants returned to the survey a second time to change their preferences (3.1% of these individuals used both methods). Note that these attempts were unsuccessful because participant randomization was determined by the first answer given by a participant, not any of the subsequent attempts made.

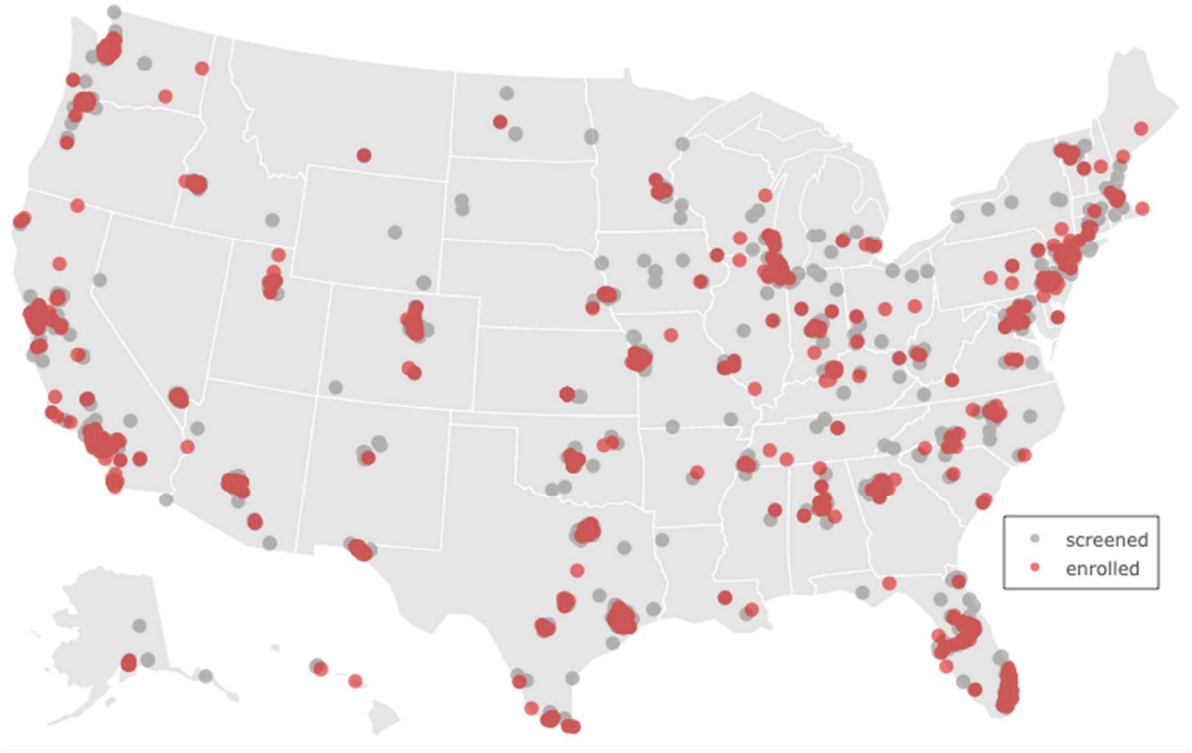


Figure 3.2. Map of US showing areas from where participants in the Brighten study were screened and enrolled.

3.4.2 *Sample Demographics*

See Table 3.1 for participant characteristics, including comparisons across those identifying as Hispanic/Latino and not. The participants were predominantly young (69.8.1% less than 40 years old; $M = 34.90$, $SD = 10.92$), female (77.1.9%), non-Hispanic White (53.3%), with 30.7% of our sample reporting Hispanic/Latino identity. The majority (69.9%) reported some form of employment, and 87.8% of all participants were iPhone users. There were significant differences between Hispanic/Latino and non-Hispanic/Latino participants; notably, a greater proportion

(40.6%) of Hispanic/Latino participants reported annual incomes of less than $20,000, compared to only 24.7% non-Hispanic/Latinos. Likewise, non-Hispanic/Latino participants were significantly more likely to be employed and more likely to have obtained a university education relative to Hispanic/Latino participants. Finally, Hispanic/Latino participants were slightly younger than their counterparts, although both groups were in their early-to-mid 30s.

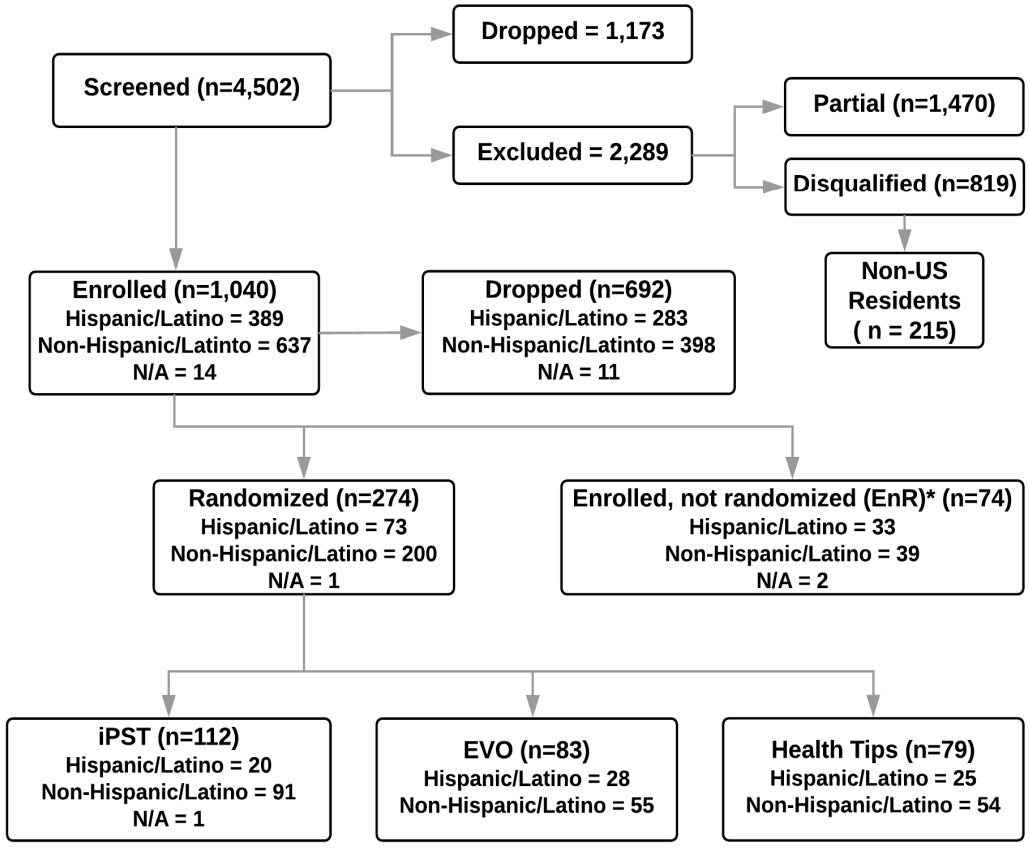


Figure 3.3. CONSORT diagram

Table 3.1. BRIGHTEN V2 participant characteristics

	Overall*	Hispanic/Latino	Non-Hispanic/Latino	*p-value*
N (%)	345	106 (30.72)	239 (69.27)	
Baseline PHQ-9 (mean (sd))	13.61 (5.46)	14.41 (5.69)	13.26 (5.34)	0.076
Gender = Female (%)	266 (77.1)	82 (77.4)	184 (77.0)	1
Age (mean (sd))	34.90 (10.92)	32.71 (10.10)	35.88 (11.15)	**0.013**
Age group (%)				0.218
18-30	137 (40.2)	51 (48.6)	86 (36.4)	
31-40	101 (29.6)	27 (25.7)	74 (31.4)	
41-50	74 (21.7)	22 (21.0)	52 (22.0)	
51-60	23 (6.7)	5 (4.8)	18 (7.6)	
61-70	5 (1.5)	0 (0.0)	5 (2.1)	
70+	1 (0.3)	0 (0.0)	1 (0.4)	
Income last year (%)				**0.005**
$20,000 or less	102 (29.6)	43 (40.6)	59 (24.7)	
20,000-40,000	90 (26.1)	31 (29.2)	59 (24.7)	
40,000-60,000	76 (22.0)	20 (18.9)	56 (23.4)	
60,000-80,000	32 (9.3)	5 (4.7)	27 (11.3)	
80,000-100,000	22 (6.4)	2 (1.9)	20 (8.4)	
100,000+	23 (6.7)	5 (4.7)	18 (7.5)	
Education (%)				**<0.001**
Community College	72 (20.9)	25 (23.6)	47 (19.7)	
Graduate Degree	58 (16.8)	11 (10.4)	47 (19.7)	
High School	56 (16.2)	29 (27.4)	27 (11.3)	
University	159 (46.1)	41 (38.7)	118 (49.4)	

Device = iPhone (%)	303 (87.8)	89 (84.0)	214 (89.5)	0.199
Working = Yes (%)	241 (69.9)	65 (61.3)	176 (73.6)	**0.03**
Race (%)				**<0.001**
Hispanic/Latinos	106 (30.7)	106 (100.0)	0 (0.0)	
Non-hispanic White	184 (53.3)	0 (0.0)	184 (77.0)	
African-American/Black	25 (7.2)	0 (0.0)	25 (10.5)	
American Indian/Alaskan Native	3 (0.9)	0 (0.0)	3 (1.3)	
Asian	24 (7.0)	0 (0.0)	24 (10.0)	
Other	3 (0.9)	0 (0.0)	3 (1.3)	
Speak Spanish = Yes (%)	113 (32.8)	96 (90.6)	17 (7.1)	**<0.001**
Income satisfaction (%)				**0.085**
Am comfortable	71 (20.6)	17 (16.0)	54 (22.6)	
Can't make ends meet	80 (23.2)	32 (30.2)	48 (20.1)	
Have enough to get along	194 (56.2)	57 (53.8)	137 (57.3)	
Marital status (%)				0.284
Married/Partner	135 (39.1)	35 (33.0)	100 (41.8)	
Separated/Widowed/Divorced	33 (9.6)	12 (11.3)	21 (8.8)	
Single	177 (51.3)	59 (55.7)	118 (49.4)	
** Participants who didn't self-report Hispanic/Latinos status (N=3) are not compared.*				

3.4.3 *Clinical Characteristics*

Overall the cohort reported moderate depressive symptomatology with a mean baseline PHQ-9 of 13.61 (*SD* = 5.46). There was no difference in baseline depression between Hispanic/Latino and non-Hispanic/Latino participants (p = .07), and neither age or gender showed a significant association with baseline PHQ-9 scores (age:[□=-0.09, p=0.06], gender:[F(1,336)=3.16, p=0.07]). Income satisfaction showed a moderate effect on baseline PHQ-9 scores (f^2= 0.265, p< 0.001). Table 3.2 summarizes the associations and effect sizes of all baseline variables with baseline PHQ-

9. Participants who reported income satisfaction as "can't make ends meet" showed significantly higher depression symptomatology (□PHQ-9 = +3.9, $p < .001$) compared to the group that reported income level as "comfortable" (Figure 3.4). However, this discrepancy in depressive symptoms between income levels was not significantly different between Hispanic/Latinos and non-Hispanic/Latinos across categories of income satisfaction.

Table 3.2. Association between demographic variables and baseline PHQ-9

Baseline variables	*Cohen's* f^2	**FDR**
Income satisfaction	0.264	< .001
Income	0.226	0.02
Spanish speaker	0.139	0.029
Education	0.160	0.076
Working	0.103	0.096
Hispanic/Latinos	0.098	0.101
Marital status	0.107	0.15
Race	0.161	0.15

3.4.4 *Cost*

Study costs beyond the initial infrastructure developed for BRIGHTEN V1 included participant payments ($7,540), website/enrollment portal/database development ($4,601), and total recruitment efforts ($14,471). A bulk of recruitment spending was for 217 Spanish language ads placed on Craigslist throughout the country ($5,725), while only $946 was spent on 33 English ads to obtain the reported enrollment. $7800 was spent on targeted social media recruitment specifically for Spanish-speakers via trialspark.com; however, only 86 unique registrants came

through this portal. Thus, participant acquisition costs differed dramatically between Spanish ($31 per enrolled participant) and English speakers ($1.49 per enrolled participant).

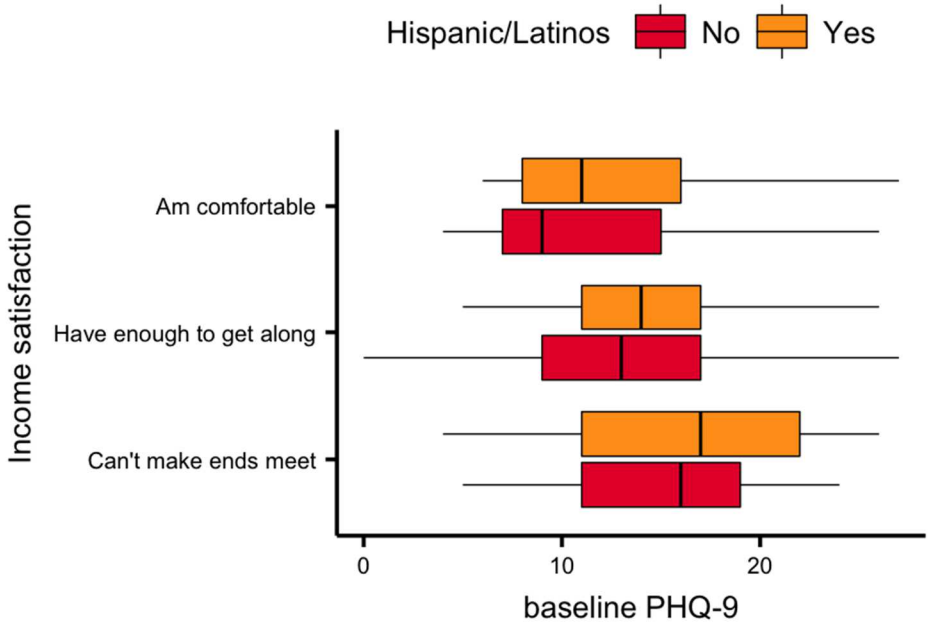


Figure 3.4. Comparison of self-reported income satisfaction and baseline depression severity.

Table 3.3. Participant acquisition costs

Recruitment Approach	Amount Spent	Number of Participants Reached	Cost per Participant
Targeted Social Media (trialspark.com for Spanish Speakers)	$7800	86	$90.70
craigslist.com (Spanish advertisements)	$5275	303	$17.41
craigslist.com (English advertisements)	$946	637	$1.49

3.4.5 *Engagement*

Overall participation in the study (as measured by assessment completion, as opposed to intervention app use) decreased by approximately 50% from week 1 to week 4, with more than 4 out of 5 participants dropping (14%) out by the end of 12 weeks. At week 4, participants contributed twice as much passive data (i.e., momentary GPS data) compared to survey assessments requiring active participation (Figure 3.5). Significant differences in participant engagement were observed between Hispanic/Latino and non-Hispanic/Latino participants (p = 0.016). Non-Hispanic/Latino individuals tended to participate in the study for 18.5 days longer than their Hispanic/Latino counterparts (*Mdn* = 53.5 days until dropout for non-Hispanic/Latinos, *Mdn* = 37 for Hispanic/Latino participants; see Figure 3.6). Finally, participants in the iPST and HTips arms were significantly more engaged compared to EVO and EnR arms ($p < .013$) regardless of the race/ethnicity (Figure 3.7).

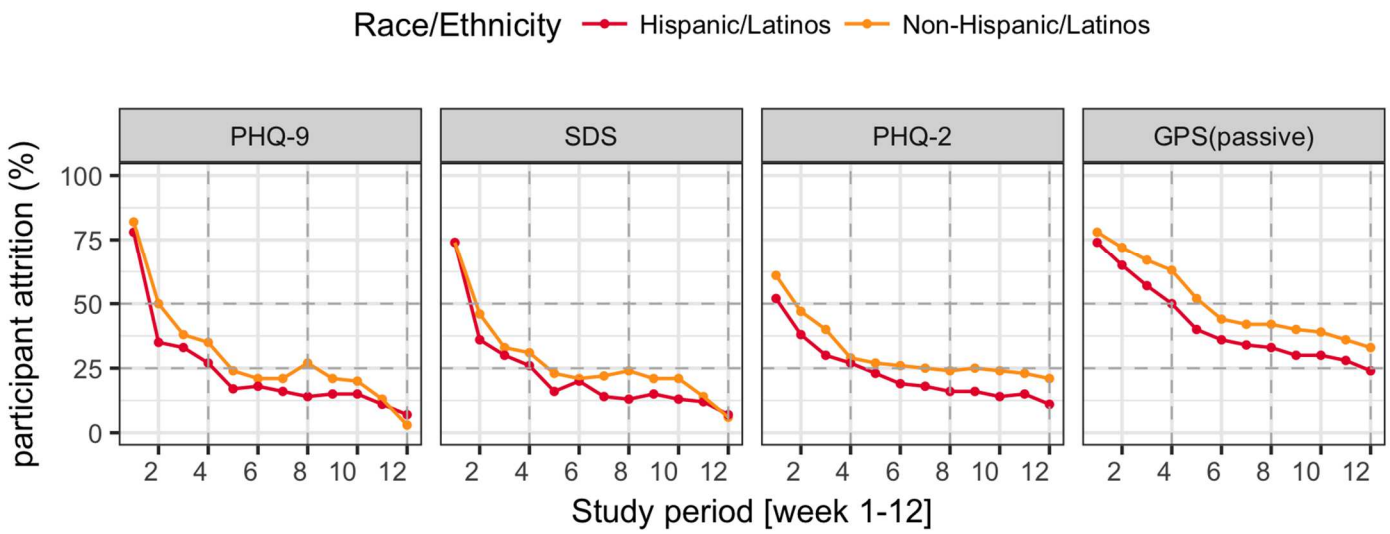


Figure 3.5. Comparison of participant attrition in the study across survey types and passive data

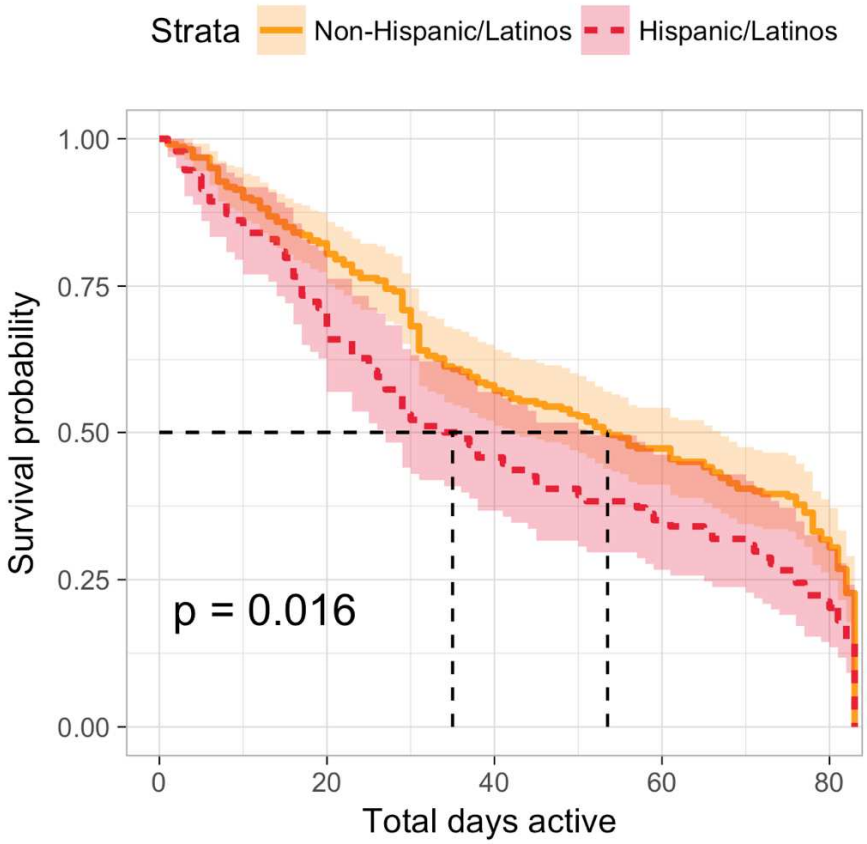


Figure 3.6. Kaplan-Meier curve comparing retention in the study across Hispanic/Latino and non-Hispanic/Latino.

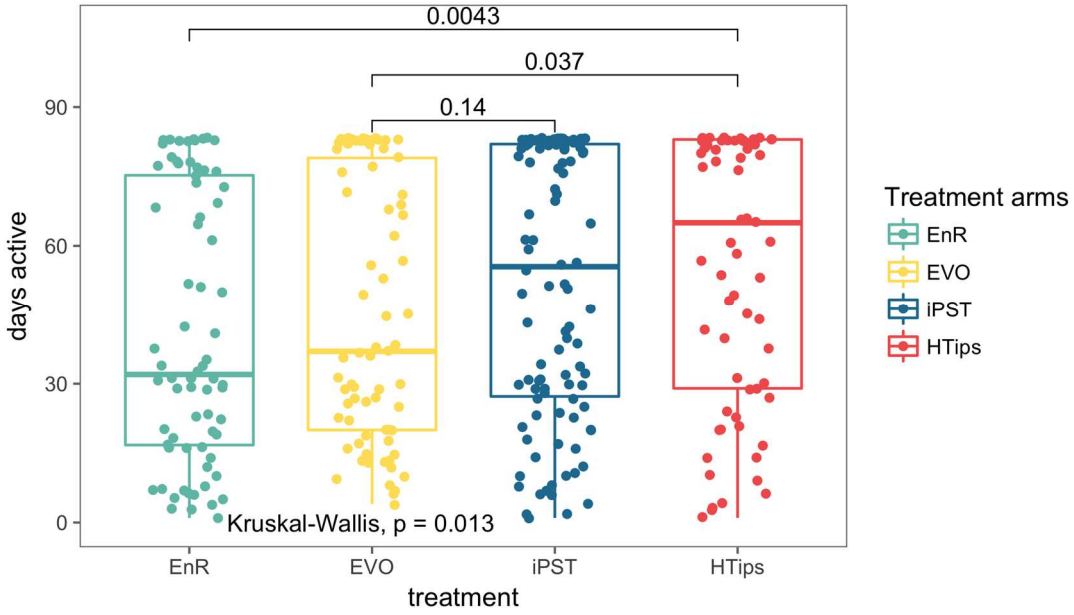


Figure 3.7. Comparison of a number of days participants were active across different treatment arms in the study.

3.4.6 *Depression Outcomes*

Change in weekly PHQ-9 scores were significantly associated with baseline severity of symptoms (i.e., mild, moderate and severe; $p < .001$). Participants who reported severe symptoms upon study entry evidenced the greatest decline in PHQ-9 scores during the first four weeks ($\beta = -4.19$, $p < .001$) but no significant further change in week 5-12. Participants with moderate symptoms also showed an initial decline in PHQ-9 ($\beta = -1.96$, $p = .004$) and further decline of 0.70 points ($\beta = -2.66$, $p = .006$) in weeks 5-12 (

Table 3.4, Figure 3.8). With regards to treatment remission at the end of week 4, 34.42% participants responded to the interventions (decrease in PHQ-9 score >= 5 from baseline), 51.63% were non-responders (change in PHQ-9 less than 5 points), and a small proportion (11.48%) deteriorated (PHQ-9 worsened >= 5 points) during the course of the study. However, there was no difference in depression outcomes between the three intervention arms. No differences in treatment remission were observed between Hispanic/Latino participants and non-Hispanic/Latinos.

3.4.7 *Disability Outcomes*

At the cohort level, disability based on SDS ratings decreased by an average 0.74 points ($p = 0.03$) in weeks 2-4 and further declined by 0.39 points ($\boldsymbol{\beta} = -1.09$, $p = .02$) in weeks 5-12. As with depression outcomes, there was no difference in disability outcomes across treatment arms. Hispanic/Latino and non-Hispanic/Latino participants did not differ in their disability outcomes (Table 3.5).

Table 3.4. Summary of estimates comparing weekly change in PHQ-9 scores using a GEE model.

	β[a]	**SE[b]**	***p-value***
Intercept	8.28	0.77	**< .001**
gender-Male	0.09	0.50	0.849
age	-0.02	0.02	0.233
week 1-4	1.33	0.55	**0.016**
week 5-12	1.33	0.72	0.064
treatment-EVO[c]	0.03	0.57	0.957
treatment-HTips[d]	-0.93	0.56	0.094
treatment-iPST[e]	-0.39	0.53	0.453
Hispanic/Latinos = Yes	-0.15	0.43	0.730

baselineState moderate	5.35	0.39	**< .001**
baselineState severe	12.26	0.46	**< .001**
week 1-4 : baselineState moderate	-1.96	0.67	**0.004**
week 5-12 : baselineState moderate	-2.66	0.96	**0.006**
week 1-4 : baselineState severe	-4.19	0.77	**< .001**
week 5-12 : baselineState severe	-4.31	1.04	**< .001**
a: Effect size, b: Standard Error, c: Project EVO, d: Health Tips, e: problem-solving therapy			

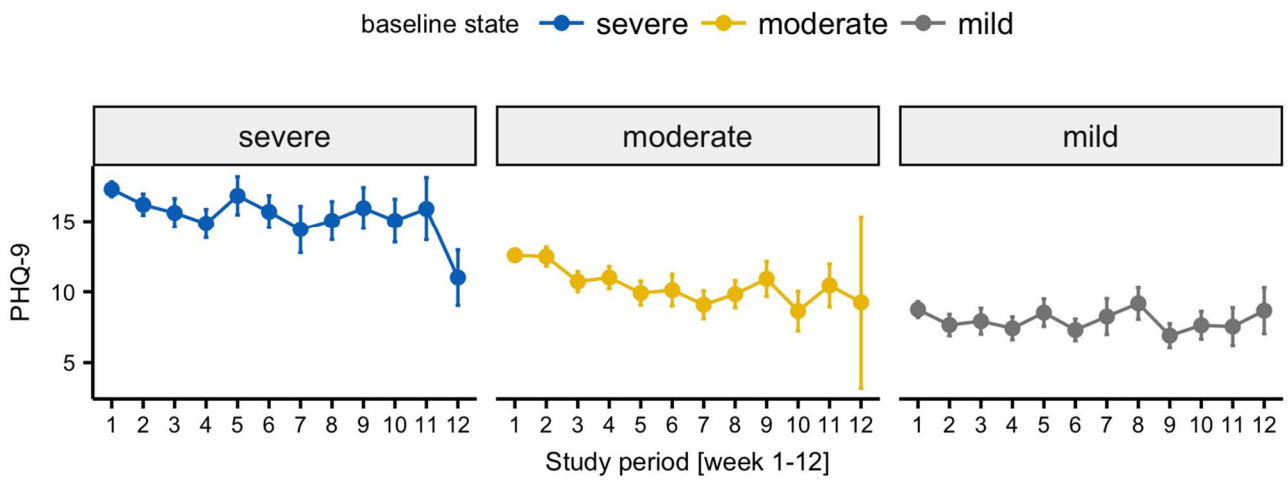


Figure 3.8. Comparison of weekly mean PHQ-9 scores with mean standard errors stratified by baseline depression state.

Table 3.5. Summary of estimates comparing weekly change in SDS score using a GEE model.

	β	SE	*p-value*
(Intercept)	10.91	1.61	0.000
genderMale	0.64	0.85	0.455
age	0.00	0.04	0.892
treatment-EVO	0.32	1.14	0.778

treatment-HTips	-0.74	1.07	0.488
treatment-iPST	-0.12	1.04	0.907
week 2-4	-0.70	0.33	**0.033**
week 5-12	-1.09	0.47	**0.019**
Hispanic/Latinos = Yes	0.12	0.82	0.881

3.5 DISCUSSION

To our knowledge, BRIGHTEN V2 is the first large-scale effort to target the remote recruitment of depressed Hispanic/Latino individuals in the United States using digital health assessments and interventions that were translated into Spanish administered solely on smartphones. We screened and enrolled one of the largest cohorts of depressed Hispanic/Latino individuals to date. Previous work has suggested that the lack of utilization of mental health care could be attributed to 1) cultural beliefs about mental health problems, 2) ineffective and inappropriate therapies, or 3) access problems or other barriers [25]. We attempted to address each of these issues by selectively targeting an underrepresented Hispanic/Latino population and using accessible, Spanish translated versions of the evidence-based intervention apps used in the initial study[9]. As has been found in other mobile-based mental health clinical trials [26,27], long-term engagement continues to be significant challenge to these studies, and this is more pronounced among Hispanic/Latinos participants. Although mobile devices are increasingly available in Hispanic/Latino communities [10], the availability of these devices as a means for conducting research and delivering care are not yet solutions that offset the widespread disparities seen in this population.

3.5.1 *Feasibility and Acceptability*

Similar to our previous work[8,28] this study has shown the feasibility of recruiting and enrolling a large and diverse sample of Hispanics/Latino adults. Previous research and observations from clinical practice suggest that Hispanics/Latino U.S. populations face barriers to research and treatment, including stigma and time constraints. This study was intended to overcome those very barriers by leveraging mobile apps that could be used at the participant's convenience. However, the engagement data showed that the Hispanics/Latino participants dropped out close to two weeks earlier than their non-Hispanics/Latinos counterparts, highlighting significant challenges in not only recruiting but keeping this population engaged. It was much more expensive and labor intensive to recruit Hispanics/Latino participants relative to the rest of the cohort. Attrition was particularly striking among the Hispanic/Latino subset, with only 18.7% (73 of 389 enrolled) downloading the treatment app. Highest dropout amongst the Hispanic/Latino sample were from participants reporting annual income level less than $20,000.

Potential issues recruiting U.S. Hispanic/Latino individuals for mental health research may hinge on (1) reluctance to be randomized, given the high number of the enrolled participants who tried to switch the initial randomly assigned intervention app, and (2) privacy concerns such as the possibility that some of our lower-income participants could be sharing the smartphones with other family members, potentially reducing the willingness to participate (hence the high initial drop out)[29]. Furthermore, the majority of participants were iPhone users, which may not be representative of the underlying population. While the ownership of an Android smartphone plus a iPad combination is relatively common as indicated by a 2014 survey[9], the ease of being able to participate in this study by only having to have a single device (iOS phone) likely spurred the bias towards iOS users in the sample.

Another potential issue in the study was the possible delay in receiving the intervention. The stratified equipoise randomization occurred after eligible participants attempted the assigned assessments (or after 72 hours, whichever came first); given that participants may have been waiting for their assigned intervention following their initial exposure to the assessment app, they may have lost interest in participating. Another consideration involves the appropriate incentive structure (e.g., timing and amount of compensation) to maximize retention and engagement, as this factor is not well understood among such underrepresented samples such as ours. It is an empirical question to understand how the amount of payment affects one's participation in a given trial. Indeed, in the first version of this study (BRIGHTEN V1), we found that participants who received bonus payment remained in the study longer than those who did not receive a bonus[9]. In that study, the experimentation of two distinct incentive models to encourage retention revealed that participant payment was not enough to keep engagement from waning. Other work has shown that externalized benefits (e.g., compensation) can dull motivation, whereas the creation of an internalized reward structure can enhance motivation and improve aspects of adherence (eg, individualized presentation of study progress, personalized encouragements) [30,31]. This is a considerable hurdle to overcome for mental health researchers who are dependent upon trying to identify features that would align with greater engagement of a culturally unique population. Thus, these issues of acceptability and engagement must be dealt with not only for research, but for any scalable intervention to take hold in routine clinical practice.

Despite the poor engagement of the active components in this study, it is clear from the findings (and those from other mobile-based studies) that there is still a tremendous potential to capture passive data from smartphone use. This form of data capture is much less burdensome as it does not require the user to actively engage with an app. If one only considers the passive data

compliance versus that of the active surveys in our study, passive data offers a viable opportunity to develop an individualized digital baseline ("digital fingerprint") and investigate deviations from baseline phone usage to behavioral fluctuations. However, using cohort-level signals in passive data to predict depression states remains modest at best[32–34], suggesting that this approach will likely require larger studies and pairing with an active task-based component for the most effective solution.

3.5.2 *Difference in Clinical Features and Outcomes*

Similar to our earlier findings in the original study[8], participants on average reported improvement in both depression and disability measures over time, regardless of treatment arm. However, more than half of the participants, regardless of their race/ethnicity, either did not evidence any clinically meaningful change or actually deteriorated according to their PHQ-9 scores. It is important to note that participants in our trial did not have a clinical diagnosis of depression; rather, they endorsed at least a mild level of depressive symptomatology at baseline screening on the PHQ-9. Moreover, treatment outcomes were based on self-report using this screening measure. Perhaps unsurprisingly, treatment response was strongest in those with greater depressive symptomatology at baseline. Thus, we interpret our clinical findings with caution, as this is not a clinical sample nor an effectiveness trial, but rather a feasibility trial in a sample of potential interest to future remote interventions. We also noted overall poor engagement in this sample with significant demographic differences between our Hispanic/Latino and non-Hispanic/Latinos participants. Hispanic/Latinos reported lower income, less income satisfaction, and lower education; such factors are previously known to be associated with an increased incidence of depression[35].

3.6 Conclusions and Future Directions

Mobile health platforms have the potential to deliver on-demand and as-needed assessment and intervention alternatives despite known barriers of time constraints, cost, stigma, and cultural and language differences. Despite the promise of closing the treatment gap for underserved communities, recruitment and retention remain problematic in such populations, and more research is needed to figure out better engagement strategies to best leverage mobile apps (e.g. appropriate incentive levels, culturally responsive content and notifications along with user-centered design approaches [36]). Like other contactless programs (e.g. self-help interventions), it is difficult to keep users engaged in active components without therapist or other in-person support [37]. However, the ubiquity and relative unobtrusive nature of smartphones does lend itself to acquiring passive sensing data, even in the absence of engagement with active components of the research or intervention protocol.

Our study offers preliminary lessons learned from doing such work in an understudied sample of Hispanic/Latinos mHealth users. Scaling these types of remote assessments and interventions will hinge on acceptance of such technology by both care teams and patients. This will be a problem for future research using remote technologies at scale to recruit and engage targeted communities (e.g., Hispanic/Latino adults with depression) and will hinge on understanding the population's needs and addressing barriers to using mental health interventions via mobile apps.

3.7 Acknowledgments

Support for this research was provided by the National Institute of Mental Health (PAA R34MH100466, T32MH0182607, K24MH074717; BNR T32MH073553) and the National Institute on Aging (JAA P30AG15272). The authors thank Thomas Egan and Tojo Chemmachel

for their help with data collection and data monitoring, Cecilia & Joaquin Anguera (author JAA parents) for their help with culturally-relevant translations within each app, website, video, and survey presented, Diana Albert for assistance in Web design, Diego Castaneda & Alinne Barrera for their willingness to speak in our promotional video and Elias Chaibub Neto for helpful insights during the data analysis phase. The authors also would especially like to thank all the participants whose time and efforts made this work possible. We would also like to thank the entire Akili Interactive team as well as Wow Labz (especially R Omanakuttan**)** for helping with data collection and partnering with us on this project.

3.8 REFERENCES

1. Mobile Fact Sheet. *Pew Research Center: Internet, Science & Tech* (2018). Available at: http://www.pewinternet.org/fact-sheet/mobile/. (Accessed: 31st January 2018)

2. Olfson, M., Blanco, C. & Marcus, S. C. Treatment of Adult Depression in the United States. *JAMA Intern. Med.* **176**, 1482 (2016).

3. Arevalo, M. *et al.* Mexican-American perspectives on participation in clinical trials: A qualitative study. *Contemp Clin Trials Commun* **4**, 52–57 (2016).

4. Miranda, J., Nakamura, R. & Bernal, G. Including ethnic minorities in mental health intervention research: a practical approach to a long-standing problem. *Cult. Med. Psychiatry* **27**, 467–486 (2003).

5. Fairburn, C. G. & Patel, V. The impact of digital technology on psychological treatments and their dissemination. *Behav. Res. Ther.* **88**, 19–25 (2017).

6. Carlbring, P., Andersson, G., Cuijpers, P., Riper, H. & Hedman-Lagerlöf, E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cogn. Behav. Ther.* **47**, 1–18 (2018).

7. Mobile Fact Sheet. *Pew Research Center: Internet, Science & Tech* (2018). Available at: http://www.pewinternet.org/fact-sheet/mobile/. (Accessed: 31st January 2018)

8. Arean, P. A. *et al.* The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. *J. Med. Internet Res.* **18**, e330 (2016).

9. Anguera, J. A., Jordan, J. T., Castaneda, D., Gazzaley, A. & Areán, P. A. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. *BMJ Innov* **2**, 14–21 (2016).

10. Dorsey, E. R. *et al.* The Use of Smartphones for Health Research. *Acad. Med.* **92**, 157–160 (2017).

11. Lavori, P. W. *et al.* Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biol. Psychiatry* **50**, 792–801 (2001).

12. Löwe, B., Unützer, J., Callahan, C. M., Perkins, A. J. & Kroenke, K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med. Care* **42**, 1194–1201 (2004).

13. Leon, A. C., Olfson, M., Portera, L., Farber, L. & Sheehan, D. V. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *Int. J. Psychiatry Med.* **27**,

93–105 (1997).

14. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).

15. Final Recommendation Statement: Depression in Adults: Screening - US Preventive Services Task Force. (1AD). Available at: https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/depression-in-adults-screening1#Pod1. (Accessed: 11th May 2018)

16. Wulsin, L., Somoza, E. & Heck, J. The Feasibility of Using the Spanish PHQ-9 to Screen for Depression in Primary Care in Honduras. *Prim. Care Companion J. Clin. Psychiatry* **4**, 191–195 (2002).

17. Sheehan, K. H. & Sheehan, D. V. Assessing treatment effects in clinical trials with the discan metric of the Sheehan Disability Scale. *Int. Clin. Psychopharmacol.* **23**, 70–83 (2008).

18. Bobes, J. *et al.* [Validation of the Spanish version of the Liebowitz social anxiety scale, social anxiety and distress scale and Sheehan disability inventory for the evaluation of social phobia]. *Med. Clin.* **112**, 530–538 (1999).

19. Anguera, J. A., Gunning, F. M. & Areán, P. A. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. *Depress. Anxiety* **34**, 508–517 (2017).

20. Liang, K.-Y. & Zeger, S. L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **73**, 13 (1986).

21. Ballinger, G. A. Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods* **7**, 127–150 (2004).

22. Rich, J. T. *et al.* A practical guide to understanding Kaplan-Meier curves. *Otolaryngol. Head Neck Surg.* **143**, 331–336 (2010).

23. Bland, J. M. The logrank test. *BMJ* **328**, 1073–1073 (2004).

24. R: The R Project for Statistical Computing. Available at: https://www.R-project.org/. (Accessed: 5th February 2018)

25. Vega, W. A., Kolody, B., Aguilar-Gaxiola, S. & Catalano, R. Gaps in service utilization by Mexican Americans with mental health problems. *Am. J. Psychiatry* **156**, 928–934 (1999).

26. Miranda, J., Azocar, F., Organista, K. C., Muñoz, R. F. & Lieberman, A. Recruiting and retaining low-income Latinos in psychotherapy research. *J. Consult. Clin. Psychol.* **64**, 868–

874 (1996).

27. Brown, G., Marshall, M., Bower, P., Woodham, A. & Waheed, W. Barriers to recruiting ethnic minorities to mental health research: a systematic review. *Int. J. Methods Psychiatr. Res.* **23**, 36–48 (2014).

28. Arean, P. A. *et al.* The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. *J. Med. Internet Res.* **18**, e330 (2016).

29. Karlson, A. K., Brush, A. J. B. & Schechter, S. Can i borrow your phone? in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (2009). doi:10.1145/1518701.1518953

30. Cruz, M., Pincus, H. A., Harman, J. S., Reynolds, C. F., 3rd & Post, E. P. Barriers to care-seeking for depressed African Americans. *Int. J. Psychiatry Med.* **38**, 71–80 (2008).

31. Van Etten, D. Psychotherapy with older adults: benefits and barriers. *J. Psychosoc. Nurs. Ment. Health Serv.* **44**, 28–33 (2006).

32. Saeb, S. *et al.* Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J. Med. Internet Res.* **17**, e175 (2015).

33. Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P. & Mohr, D. C. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ* **4**, e2537 (2016).

34. Pratap, A. *et al.* The feasibility of using smartphones to assess and remediate depression in Hispanic/Latino individuals nationally. in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers on - UbiComp '17* (2017). doi:10.1145/3123024.3127877

35. Lorant, V. Socioeconomic Inequalities in Depression: A Meta-Analysis. *Am. J. Epidemiol.* **157**, 98–112 (2003).

36. Vredenburg, K., Mao, J.-Y., Smith, P. W. & Carey, T. A survey of user-centered design practice. in *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02* (2002). doi:10.1145/503457.503460

37. Aguilera, A. Digital Technology and Mental Health Interventions: Opportunities and Challenges. *Arbor* **191**, a210 (2015).

Chapter 4. PARTICIPANT ENROLLMENT AND RETENTION IN REMOTE DIGITAL HEALTH STUDIES

4.1 ABSTRACT

Digital technologies such as smartphones are transforming the way scientists conduct biomedical research using real-world data. Several remotely conducted studies have recruited thousands of participants over a span of a few months. Unfortunately, these studies are hampered by substantial participant attrition, calling into question the representativeness of the collected data including generalizability of findings from these studies. We report the challenges in retention and recruitment in eight remote digital health studies comprising over 100,000 participants who participated for more than 850,000 days, completing close to 3.5 million remote health evaluations. Survival modeling surfaced several factors significantly associated($P < 1e\text{-}16$) with increase in retention time **i)** Clinician referral(an increase of 40 days in median retention time), **ii)** Effect of compensation (22 days), **iii)** Clinical conditions of interest to the study (7 days) and **iv)** Older adults(4 days). Additionally, four distinct patterns of daily app usage behavior that were also associated ($P < 1e\text{-}10$) with participant demographics were identified. Most studies were not able to recruit a representative sample, either demographically or regionally. Combined together these findings can help inform recruitment and retention strategies to enable equitable participation of populations in future digital health research.

4.2 INTRODUCTION

Traditional in-person clinical trials serve as the cornerstone of modern healthcare advancement. While a pivotal source of evidence generation for advancing clinical knowledge, in-person trials

are also costly and time-consuming, typically running for 3-5 years from conception to completion, at a cost of millions of dollars per study. These timelines have often meant that promising treatments take years to get to market, which can create unnecessary delays in advancing clinical practice. Additionally, clinical research suffers from several other challenges[1,2] including 1) recruiting sufficiently large and diverse cohorts quickly, and 2) tracking day-to-day fluctuations in disease severity that often go undetected in episodic in-clinic evaluations[3,4]. Scientists have recently turned to digital technology[5,6] to address these challenges, hoping to collect real-world evidence[7] from large and diverse populations to track long-term health outcomes and variations in disease trajectories at a fraction of the cost of traditional research[8].

The global penetration[9] and high-frequency usage of smartphones (up to 4 hours daily[10,11]) offer researchers a cost-effective means to recruit a large number of participants into health research across the US (and the world)[12,13]. In the last 5 years, researchers have conducted several large scale studies[14–22] including deploying interventions[23,24] and running clinical trials[25–27] using mobile technologies. These studies are able to recruit at-scale because participants can be identified and consented[28] to participate in the study without ever having stepped foot in a research lab, with significantly lower costs than conventional clinical trials[23,24]. Mobile technologies also allow investigators an opportunity to collect data in real-time based on people's daily lived experiences of the disease, that is, real-world data[7]. Rather than retrospectively asking people to recall their health over the past week or month, researchers using mobile technologies can assess participants frequently including outside clinic and at important points in time without having to rely on recall that is known to have bias[29]. While these studies show the utility of mobile technology, challenges in participant diversity and long-term participant retention still remain a problem[30].

Digital studies continue to suffer from long-term participant retention problems that also plagued internet-based studies[31,32] in the early 2000s[33–35]. However, our understanding of factors impacting retention in remote research remains to be limited. High levels of user attrition combined with variations in long-term app usage may result in the creation of a cohort that may not represent the population of interest in regard to demographics, disease status, and disability. This has called into question the reliability and utility of the collected data from these studies[36]. Furthermore, while for many digital health studies, anyone eligible can self-select to join, this broad "open enrollment" recruitment model may be prone to selection and ascertainment bias[36]. Systematic evaluation of participant recruitment and retention could help detect such confounding characteristics that may be present in large scale remotely collected data and has been shown to severely impact the generalizability of the derived statistical inference[36,37]. Participant retention may also be partially dependent on the engagement strategies used in remote research. While most studies assume participants will remain in a study for altruistic reasons[38], other studies provide compensation for participant time[39], leverage partnerships with local community organizations, clinical registries, and clinicians to encourage participation[23,24]. Although monetary incentives are known to increase participation in research[40], we know little about the relative impact of demographics, recruitment and different engagement strategies on participant retention, especially in remote health research.

The purpose of this study is to document the drivers of retention, and long-term study app usage in remote research. To investigate these questions, we have compiled user engagement data from eight digital health studies that enrolled more than 100,000 participants from throughout the US between 2014-2019. These studies assessed different disease areas including asthma, endometriosis, heart disease, depression, sleep health, neurological diseases and consisted of a

combination of longitudinal subjective surveys and objective sensor-based tasks including passive data[41] collection. The diversity of the collected data allows for a broad investigation of different participant characteristics and engagement strategies that may be associated with higher retention including assessment of representational bias in the collected real-world data.

4.3 METHODS

4.3.1 *Data Acquisition*

The user engagement data was collected from eight digital health studies assessing different diseases ranging from parkinson's, asthma, heart condition, sleep health, multiple sclerosis to depression (Table 4.1). These studies recruited participants from throughout the US between 2014-2019 using a combination of different approaches including placing ads on social media, publicizing or launching the study at a large gathering, partnerships with patient advocacy groups, clinics, and through word of mouth. The studies were launched at different time points during the 2014-2019 period, including three studies mPower, MyHeartCounts, and Asthma being launched with the public release of ResearchKit framework[73] released by Apple in March 2015. The studies were also active for different time periods including significant differences in the minimum time participants were expected to participate in the studies remotely. While Brighten and ElevateMS had a fixed 12 week participation period, other studies allowed participants to remain active for as long as they desired. Given this variation in the expected participation period across the studies, we selected the minimum common time period of the first 12 weeks(84 days) of each participant's activity in each study for retention analysis. Finally, with the exception of Brighten study which was a randomized interventional clinical trial and enrolled depressed cohort offering them monetary incentives for participation, the rest of the seven studies were observational and did not

offer any direct incentives for participation and were open to people with and without target disease. The studies also collected different real-world data ranging from frequent subjective assessments, objective sensor-based tasks to continual passive data[41] collection.

4.3.2 *Data Harmonization*

User activity data across all the apps were harmonized to allow for inter-app comparison of user engagement metrics. All in-app surveys and sensor-based tasks (eg. Finger tapping on the screen) were classified as "active tasks" data type. The data gathered without explicit user action such as daily step count (Apple's health kit API), daily local weather patterns were classified as "passive" data type and was not used for assessing active user engagement. The frequency at which the active tasks were administered in the study apps were aligned based on the information available in the corresponding study publication or obtained directly from the data contributing team in case the data was not publicly available. Furthermore, there were significant differences in the baseline demographics that were collected by each app. A minimal subset of four demographic characteristics (age, gender, race, state) was used for participant recruitment and retention analysis. A subset of six studies(mPower, ElevateMS, SleepHealth, Asthma, MyHeartCounts) had enrolled participants with and without disease status and were used to asses retention differences between people with(case) and without(control) disease. Two studies (mPower and ElevateMS) had a subset of participants that were referred to use the same study app by their care providers. For this smaller but unique sub-group, we compared the retention differences between clinically referred participants to self-referred participants.

4.3.3 *Statistical Analysis*

We used three key metrics to assess participant retention and long-term engagement. 1) Duration in the study: the total duration, a study participant remained active in the study i.e the number of days between the first and last active task completed by the participant, during the first 84 days of each participant's time in the study. 2) Days active in the study: the number of days a participant performed any active task in the app. 3) User activity streak: a binary-encoded vector representing the 84 days of app participation for each participant (Figure 4.3-a) where the position of the vector indicates the participant's day in the study and is set to 1 (green box, Figure 4.3-a) if at least one active task is performed on that day or else is 0 (white). User activity streak was used to assess sub-populations that show similar longitudinal engagement patterns over a 3-month period.

Participant retention analysis(survival analysis[74]) was done using the total duration in the study metric to compare the retention differences across studies, sex, age group, disease status, and clinical referral for study-app usage. Log-rank test[75] stratified by study type was used to compare significant differences in participant retention between different comparator groups. Kaplan-Meier[76] plots were used to summarize the effect of the main variable of interest by pooling the data across studies where applicable. Two approaches were used to evaluate participant retention using survival analysis. 1) No censoring (most conservative) - If the last active task completed by participant fell within the pre-specified study period of first 84 days, we considered it to be a true event i.e participant leaving the study (considered "dead" for survival analysis). b) Right-censoring[76] - To assess the sensitivity of our findings using approach 1, we relaxed the determination of true event (participant leaving the study) in the first 12 weeks to be based on the first 20 weeks of app activity (additional 8 weeks). For example, if a participant completes last task in an app on day 40 (within the first 84 days) and then additionally completes more active

task/s between week 13-20 he/she was still considered alive (no event) during the first 84 days (12 weeks) of the study and therefore "right-censored" for survival analysis.

Given that age and gender had a varying degree of missingness across studies; additional analysis comparing the retention differences between the two sub-groups that provided the demographics and that opted out was done to assess the sensitivity of missing data on main findings. Unsupervised k-means clustering was used to investigate the longitudinal participant engagement behavior within each study. The number of optimum clusters(between 1-10) in each study was determined using the elbow method[77] that aims to minimize the within-cluster variation. Enrichment of demographic characteristics in each cluster was assessed using a one-way analysis of variance (ANOVA). Since the goal of this unsupervised clustering of user activity streaks was to investigate the patterns in longitudinal participant engagement; we filtered out individuals who remained in the study for less than 7 days from clustering analysis. However, for post hoc comparisons of demographics across the clusters, the initially left-out users were put in a separate group (C5*). The state-wise proportions of recruited participants in each app were compared to the 2018 US state population estimates using the data obtained from the US census bureau[78]. To eliminate potential bias related to marketing and advertising of the launch of Apple's Research kit platform on March 09, 2015, participants who joined and left the mPower, MyHeartCounts, Asthma studies within the first week of Research Kit launch (N=15,413) were taken out from the user retention analysis. We initially considered using cox proportional hazards model[79] to test for the significance of variable of interest on user retention within each study accounting for other study-specific covariates. However, because the assumption of proportional hazards (tested using scaled Schoenfeld residuals) was not supported for some studies, these analyses were not further pursued. All statistical analyses were performed using R[80].

Table 4.1. Summary of user engagement data compiled from eight digital health studies

Study	Disease Focus / Study type	Study period	Number of participants	Total participant days	Active tasks completed
Start	Antidepressant Efficacy - Observational	Aug,2015 - Feb,2018	42,704	280,489	1,219,656
MyHeartCounts	Cardiovascular Health - Observational	Mar,2015 - Oct,2015	26,902	165,455	305,821
SleepHealth	Sleep Apnea - Observational	Jul,2015 - Jun,2019	12,914	99,696	401,628
mPower	Parkinson's - Observational	Mar,2015 - Jun,2019	12,236	104,797	568,685
Phendo	Endometriosis - Observational	Dec,2016 - Jul,2019	7,802	81,938	735,778
Asthma	Asthma - Observational	Mar,2015 - Dec,2016	5,875	77,815	175,699
Brighten	Depression - Randomized Control Trial	Jul,2014 - Aug,2015	876	34,987	45,951
ElevateMS	Multiple Sclerosis - Observational	Aug,2017 - Jul,2019	605	11,211	31,568
			109,914	**856,388**	**3,484,786**

4.4 RESULTS

4.4.1 *Participant Characteristics*

The combined user activity data from eight digital health studies resulted in a pool of 109,914 participants who together completed approximately 3.5 million tasks on more than 850,000 days (Table 4.1). Across the studies, the majority (Median=65.2%) of participants were between 17-40 years with those 60 years and older being the least represented (Median across studies=6% of the study population). The sample had a larger proportion of Females (Median=56.9%) however it varied significantly across the studies (Range=29.4-100%). A majority of recruited participants were Non-Hispanic Whites (Median=75.3%) followed by Hispanic/Latinos (Median=8.21%) and African-American/Blacks (Median=3.45%) (Table 4.2). With the exception of the Brighten study, the race/ethnic diversity of the sample also showed a marked difference from the 2010 census data.

Minority groups were under-represented in the present sample with Hispanic/Latinos and African-America/Black showing a substantial difference of -8.09% and -9.15% respectively compared to the 2010 census metrics (Table 4.2, Figure 4.1-b). Across the studies, the median proportion of recruited participants per state showed notable differences from the state's population proportion of the US (Figure 4.1-a).

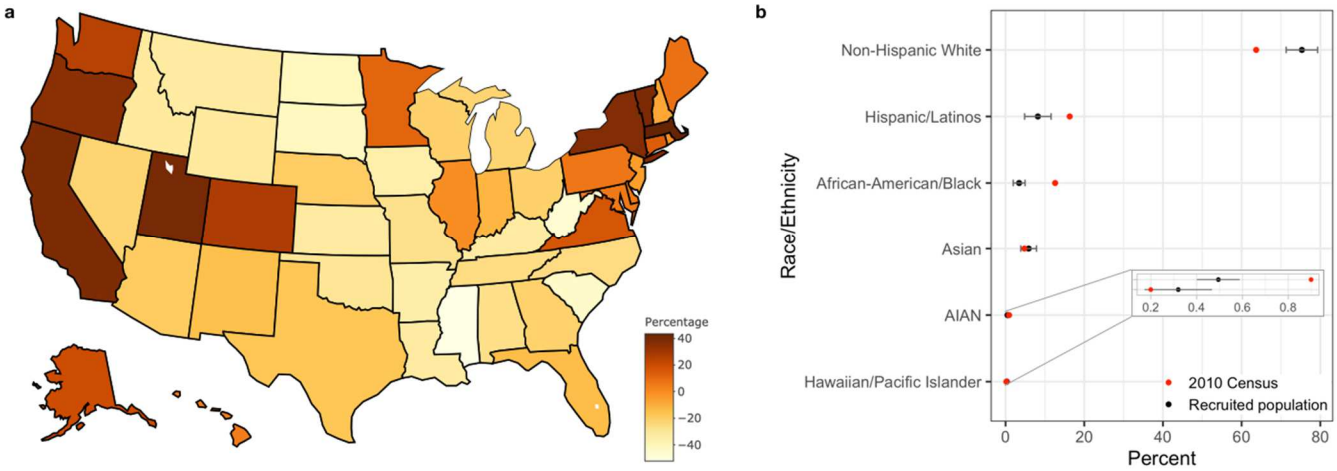


Figure 4.1**.** Comparison of geographical and race/ethnic diversity of the study sample to general US population.

Map of US showing the ratio of the percentage of recruited participants to state's population proportion of the US (median across the studies) and **b)** Race/Ethnicity proportion (median +/- IQR) of recruited participants compared to 2010 census data.

4.4.2 *Participant Retention*

As is the nature of these studies, participants were required to complete all health assessments and other study-related tasks (eg: treatments) through a mobile application (app) throughout the length of the study. The median time participants engaged in the study in the first 12 weeks was 5.5 days

of which in-app tasks were performed on 2 days (Table 4.2). Higher proportions of active tasks were completed by participants during the evening (4-8 PM) and night (8-12 Midnight) hours (Figure 4.2-a). Across the studies, the median retention time varied significantly (P <1e-16) between 2 and 12 days with the Brighten study being an outlier with a higher median retention of 26 days (Figure 4.2-b). A notable increase in median retention time was seen for sub-cohorts that continue to engage with the study apps after day one and beyond (Figure 4.2-c). For example, the median retention increased by 25 days for the sub-cohort that was engaged for the first 8 days. The participant retention also showed a significant association with participant characteristics. While older participants (60 years and above) were the smallest proportion of the sample, they remained in the study for a significantly longer duration (Median=7 days, P<1e-16) compared to the majority younger sample (17-49 years) (Figure 4.2-d). Participants declared gender showed no significant difference in retention (P = 0.3). People with clinical conditions of interest to the study (e.g.: heart disease, depression, multiple sclerosis) remained in the studies for a significantly longer time (Median=13 days, P<1e-16) compared to participants that were recruited as non-disease controls (Median=6 days) (Figure 4.2-e). Median retention time also showed a marked and significant increase of 40 days (P<1e-16) for participants that were referred by a clinician to join one of the two studies (mPower and ElevateMS)(Median=44 days) compared to participants who self-selected to join the same study (Median=4 days) (Figure 4.2-f). See Supplementary Tables 1-6(Appendix A) for a further breakdown of survival analysis results. Sensitivity analysis by including participants with missing age showed no impact on the association of age with participant retention. However, participants with missing demographics showed variation in retention compared to participants who shared their demographics (Supplementary Figure 1,

Appendix A). This could be related to different time points at which demographic related questions were administered in individual studies.

Table 4.2**.** Summary of select participant demographics and study app usage across the eight digital health studies

	Asthma	**Brighten**	**ElevateMS**	**mPower**	**MyHeart Counts**	**Phendo**	**SleepHealth**	**Start**	**Overall (median)**
Age group									
N	*2512*	*875*	*569*	*6810*	*1555*	*7484*	*12392*	*42690*	
18-29 (%)	43.31	50.06	10.9	31.5	25.08	55.38	32.79	55.72	38
30-39 (%)	27.83	25.14	26.54	18.37	32.67	36.09	28.72	24.14	27.2
40-49 (%)	14.41	14.74	28.47	13.19	16.27	8.23	20.77	12.38	14.6
50-59 (%)	9.08	6.97	22.14	13.61	12.09	0.25	11	5.26	10
60+ (%)	5.37	3.09	11.95	23.33	13.89	0.04	6.72	2.51	6
Sex									
N	*2509*	*875*	*329*	*6916*	*6976*	*7532*	*12558*	*42704*	
Female (%)	39.58	77.83	74.16	28.93	18.94	100	29.14	75.86	56.9
Race									
N	*3274*	*875*	*334*	*6884*	*4703*	*7530*	*5311*	-	
Non-Hispanic White (%)	68.69	60.11	80.84	75.32	77.95	81.29	74.13	-	75.3
Hispanic/Latinos (%)	13.29	14.29	4.79	8.21	6.97	5.67	12.82	-	8.21
African-American/Black (%)	4.95	10.86	6.89	2.05	3.1	2.71	3.45	-	3.45
Asian (%)	4.98	8.23	2.99	8.4	7.72	2.79	5.87	-	5.9
Hawaiian or other Pacific Islander (%)	0.89	0.57	0	0.28	0.32	0	0.23	-	0.3
AIAN (%)	0.43	0.46	0	0.65	0.53	0.74	0.28	-	0.5
Other (%)	6.78	5.49	4.49	5.1	3.4	6.8	3.22	-	5.1
Duration in Study (Median +/- IQR)	12 ± 38	26 ± 82	7 ± 45	4 ± 21	9 ± 19	4 ± 25	2 ± 8	2 ± 16	5.5
Days active tasks performed (Median +/- IQR)	4 ± 12	14 ± 58	2 ± 8	2 ± 4	4 ± 7	2 ± 6	2 ± 4	2 ± 4	2

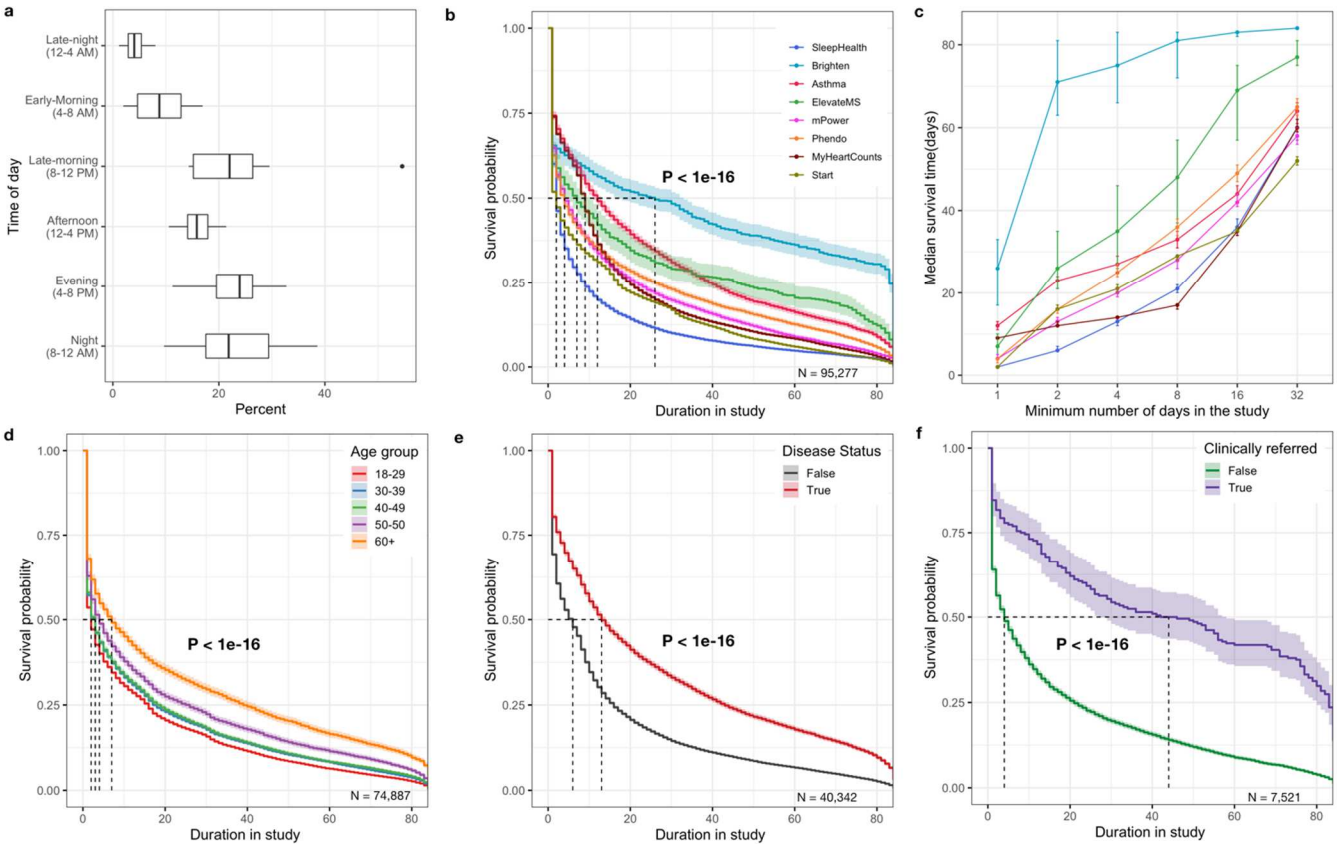


Figure 4.2. Kaplan Meir survival curves comparing retention differences across participant characteristics

a) Proportion of active tasks (N = 3.3 million) completed by participants based on their local time of day, **b)** Kaplan Meir survival curve showing significant differences (P < 1e-16) in user retention across the apps. Brighten App where monetary incentives were given to participants showed the longest retention time(Median = 26 days, 95% CI= 17-33) followed by Asthma(Median = 12 days, 95% CI= 11-13), MyHeartCounts(Median = 9 days, 95% CI= 9-9), ElevateMS(Median = 7 days, 95% CI= 5-10), mPower(Median = 5 days, 95% CI= 4-5), Phendo(Median = 4 days, 95% CI= 3-4), Start(Median = 2 days, 95% CI= 2-2) and SleepHealth(Median = 2 days, 95% CI= 2-2), **c)** Lift curve showing the change in median survival time (with 95% CI) based on the minimum number of days(1-32) a subset of participants continued to use the study apps, Kaplan-Meier survival curve showing significant differences in user retention across **d)** Age group, with 60 years and older using the apps for longest duration(Median=7days, 95%CI=6-8, P < 1e-16) followed by 50-59 years (Median =4 days, 95%CI= 4-5) and 17-49 years (Median = 2-3 days, 95% CI= 2-3) **e)** Disease status; participants reporting having a disease stayed active longer(N_{50}= 13days, 95% CI=13-14) compared to people without disease(N_{50}= 6 days, 95% CI=5-6) and finally **f)** Clinical referral; Two studies (mPower and ElevateMS), had a subpopulation, that were referred to the study by clinicians and showed significantly(P<1e-16) longer

app usage period(Median= 44 days, 95% CI=27-58) compared to self-referred participants with disease (N_{50}= 4 days, 95% CI=4-4).

4.4.3 *Participant Daily Engagement Patterns*

In the subgroup of participants who engaged with study apps for a minimum of 7 days, overall app usage clustered into four distinct groups with high (the dedicated cluster C1, and high utilizers in C2), moderate (cluster C3) and Sporadic (cluster C4) engagement (Figure 4.3-b). The participants who did not participate for at least 7 days were placed in a separate group of participants (the abandoners-C5*) (See Methods for cluster size determination and exclusion criteria details). The engagement and demographic characteristics across these five groups (C1-5*) varied significantly. Cluster 1 and 2 showed the highest daily app usage (Median app usage in the first 84 days = 96.4% and 63.1 % respectively) but also had the smallest proportion of participants (Median =9.5%) with the exception of Brighten where 23.7% of participants belonged to cluster C1. While daily app usage declined significantly for both moderate and sporadic clusters (C3- 21.4% and C4-22.6%), the median number of days between app usage was significantly higher for participants in the sporadic C4 cluster (Median=5 days) compared to cluster C3(Median=2 days). The majority of participants (median 54.6%) across the apps were linked to the abandoner group(C5*) with the median app usage of just 1 day (Figure 4.4 a-b). Furthermore, distinct demographic characteristics emerged across these five groups. Higher engagement clusters (C1-2) showed significant differences (P=1.38e-12) in proportion of adults 60 years and above (Median range =15.1-17.2% across studies) compared to lower engagement clusters C3-5*(Median range =5.1-11.7% across studies) [Figure 4.4-c]. Minority groups such as Hispanic/Latinos, Asians, and African-American/Black, on the other hand, were represented in higher proportions in the clusters (C3-5*)

(P=4.12e-10) with the least engagement (Figure 4.4-d) (See Supplementary Table 8, Appendix A for further details).

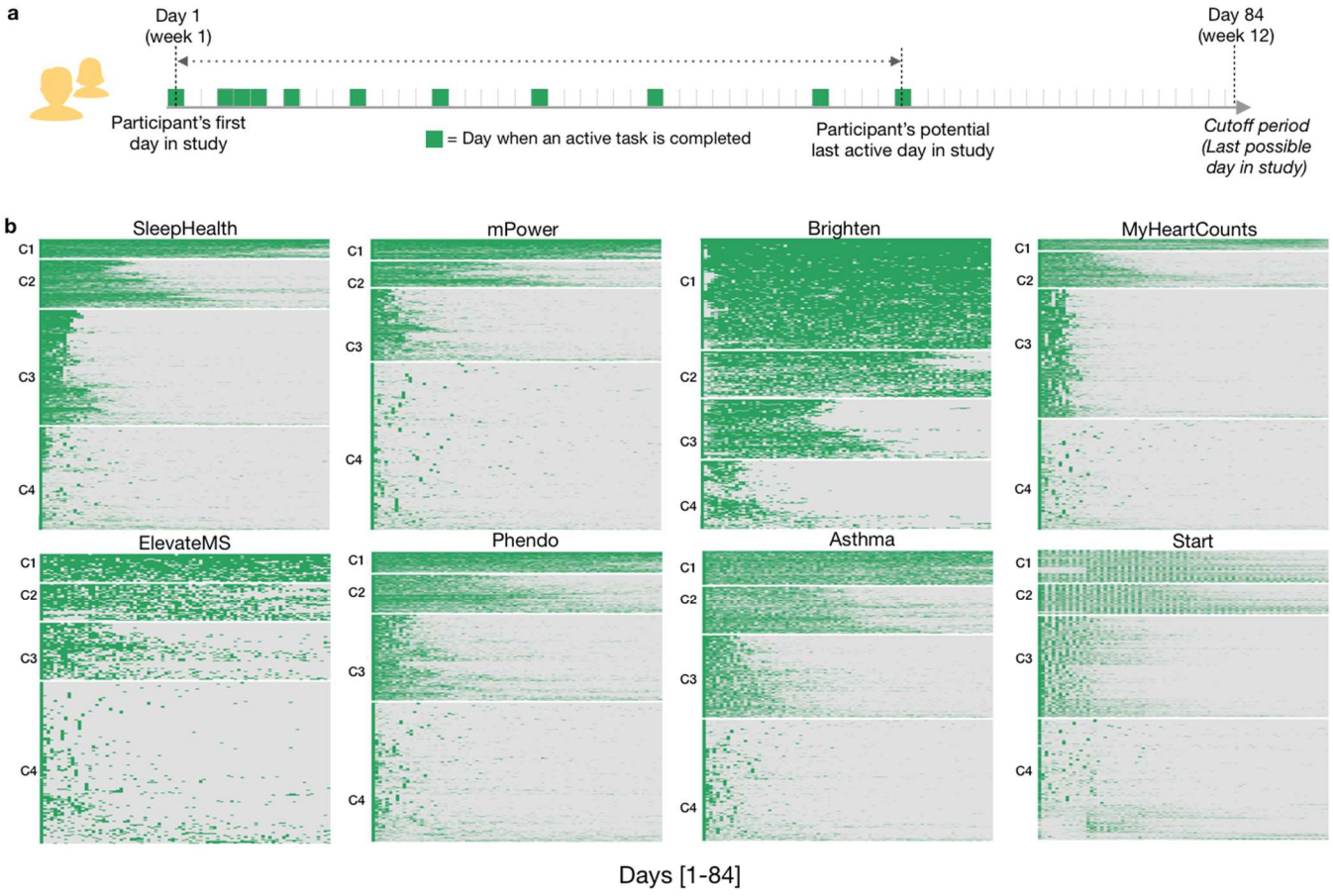


Figure 4.3. Comparing trends in long term app usage

a) Schematic representation of an individual's in-app activity for the first 84 days. The participant app usage time is determined based on the number of days between the first and last day they perform an active task (indicated by the green box) in the app. Days active in the study is the total number of days a participant performs at least one active task (indicated by the number of green boxes). **b)** Heatmaps showing participants in-app activity across the apps for the first 12 weeks (84 days), grouped into four broad clusters using unsupervised k-means clustering. The optimum number of clusters was determined by minimizing the within-cluster variation across different cluster sizes between 1-10. Seven out of eight studies indicated four clusters to be an optimum

number using the elbow method. The heatmaps are arranged by the highest (C1) to the lowest user engagement cluster (C4).

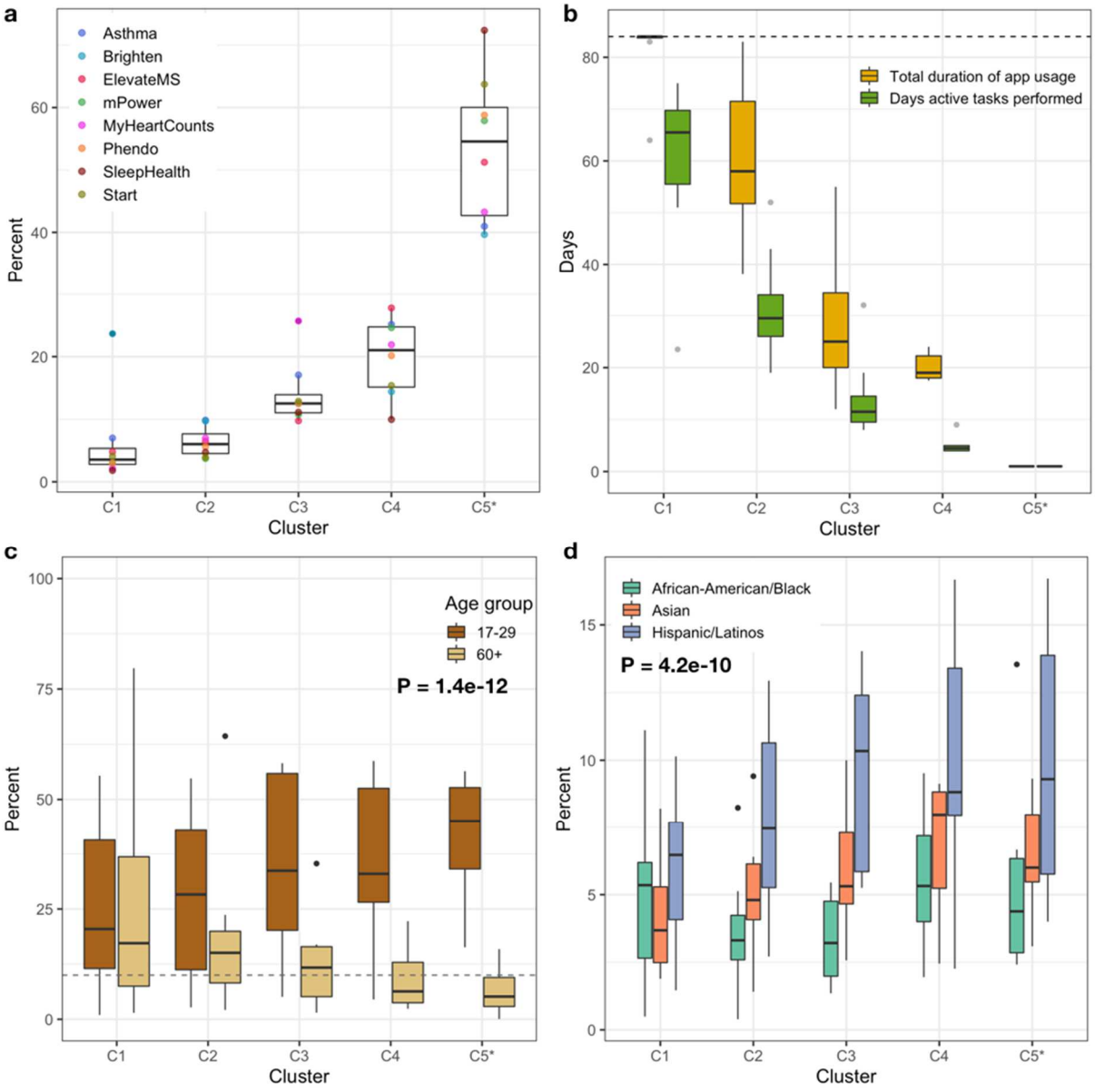


Figure 4.4. Comparison of characteristics across five long term app usage clusters **a)** Proportion of participants in each cluster across the study apps, **b)** Participants total app usage duration(between 1-84 days) and the number of days participants completed tasks in the study apps, **c)** Significant differences[F(4,163)=18.5, P=1.38e-12] in proportion of participants aged 17-29 years

and 60 years and older across the 5 clusters and **d)** Significant differences[F(2,81)=28.5, P=4.12e-10] in proportion of minority population present in the five clusters. C5* cluster contains the participants that used the apps for less than a week and were removed from the clustering, however, they were added back for accurate proportional comparison of participants in each cluster.

4.5 DISCUSSION

Our findings are based on one of the largest and diverse engagement data set compiled to date. We identified two major challenges with remote data collection: (1) more than half of the participants discontinued participation within the first week of a study but that the rates at which people discontinued was drastically different based on age, disease status, clinical referral, and use of monetary incentives and (2) most studies were not able to recruit a representative sample, either demographically or regionally. Although these findings raise questions about the reliability and validity of data collected in this manner, they also shed light on potential solutions to overcome biases in populations using a combination of different recruitment and engagement strategies.

One solution could be the use of a flexible randomized withdrawal design[42]. Temporal retention analysis (Figure 4.2-c) shows that a run-in period could be introduced in the research design, wherein participants who are not active in the study app in the first week or two of the study can be excluded after enrollment but before the start of the actual study. The resulting smaller but more engaged cohort will help increase the statistical power of the study but does not fix the potential bias[43].

Another solution is to rely on monetary incentives to enhance engagement. Although only one study paid participants, the significant increase in retention and the largest proportion of frequent app users indicate the utility of the fair-share compensation model[1,44,45] in remote research. Such "pay-for-participation" model could be utilized by studies that require long-term and frequent

remote participation. Researchers conducting case-control studies should also plan to further enrich and engage the population without the disease. Studies run the risk of not collecting sufficient data from controls to perform case-control analysis with participants without disease seen to be dropping out significantly early. Similarly, more efforts[46–48] are needed to retain the younger population that, although demonstrating large enrollment also features a majority dropping out on day one.

Distinct patterns in daily app usage behavior, also shown previously[49], further strengthen the evidence of unequal technology utilization in remote research. The majority of the participants found in the abandoners group (C5*) who dropped out of the study on day 1, may also reflect initial patterns in willingness to participate in research, in a way that cannot be captured by recruitment in traditional research. Put another way, although there is significant dropout in remote trials, these early drop-outs may be able to yield very useful information about differences in people who are willing to participate in research and those who are not willing to participate. For decades clinical research has been criticized for its potential bias because people who participate in research may be very different from people who do not participate in research[50–52]. Although researchers will not have longitudinal data from those who discontinue participation early, the information collected during onboarding can be used to assess potential biases in the final sample and may inform future targeted retention strategies.

Only 1 in 10 participants were in the high app use clusters (C1-2), and these clusters tended to be largely Non-Hispanic whites and older adults. Minority and younger populations, on the other hand, were represented more in the clusters with the lowest daily app usage (Figure 4.4-d). The largest impact on participant retention (>10 times) in the present sample was associated with clinician referral for participating in a remote study. This referral can be very light touch in nature,

for example in the ElevateMS study, it consisted solely of clinicians handing patients a flyer with information about the study during a regular clinic visit. This finding is understandable, given recent research[53] showing that the majority of Americans trust medical doctors.

With the exception of Brighten study, the recruited sample was also inadequately diverse highlighting a persistent digital divide[54] and continued challenges in the recruitment of racial and ethnic underserved communities[55]. Additionally, the underrepresentation of States in the southern, rural and midwest regions indicates that areas of the US that often bear a disproportionate burden of disease[56] are under-represented in digital research[57,56,58]. This recruitment bias could impact future studies that aim to collect data for health conditions that are more prevalent among certain demographic[59] and associated with geographic groups[60]. Using different recruitment strategies[46–48] including targeted online ads in regions known to have a larger proportion of the minority groups, partnerships with local community organizations and clinics may help improve the penetration of remote research and improve diversity in the recruited sample. The ongoing "All of Us" research program that includes remote digital data collection has shown the feasibility of using a multifaceted approach to recruit a diverse sample with a majority of the cohort coming from communities underrepresented in biomedical research[61]. Additionally, simple techniques such as stratified recruitment that is customized based on the continual monitoring of the enrolling cohort demographics, can help enrich for a target population.

Finally, communication in digital health research may benefit from adopting the diffusion of innovations approach[62,63] that has been applied successfully in healthcare settings to change behavior including the adoption of new technologies[64–66]. Research study enrollments, advertisements including in-app communication and return of information to participants[67], could be tailored to fit three distinct personality types (trendsetters, majority, and laggards). While

trendsetters will adopt innovations early, they are a minority (15%) compared to the majority (greater than two-thirds of the population) who will adopt a new behavior after hearing about its real-benefits, utility and believe it is the status quo. On the other hand, laggards (15%) are highly resistant to change and hard to reach online and as a result, will require more targeted and local outreach efforts.

These results should also be viewed within the context of limitations related to integrating diverse user-engagement data across digital health studies that targeted different disease areas with varying underlying disease characteristics and severity. We did not take into account differences in recruitment strategies used by the study apps. The present retention analysis is based on the "completed" tasks and did not account for incomplete tasks or time participants spent in the app. While sensitivity analysis showed the main findings from user retention analysis do not change by including participants with missing data, however, missing demographic characteristics remains to be a significant challenge for digital health (See Supplementary Table 7, Appendix-A). Researchers should prioritize to collect minimal demographic data such as age, gender, race/ethnicity, participant state during onboarding which help characterize user attrition in future studies.

Despite these limitations, the present investigation to the best of our knowledge is the largest cross-study analysis of participant retention in remote digital health studies. While the technology has enabled researchers to reach and recruit participants for conducting large scale health research in short periods of time, more needs to be done to ensure equitable access and long-term utilization by participants across different populations. The low retention in "fully remote, app-based" health research may also need to be seen in the broad context of the mobile app industry where similar user attrition is reported[68]. Attrition in remote research may also be impacted by study burden[30]

as frequent remote assessments can compete with users' everyday priorities and perceived value proposition for completing a study task that may not be linked to an immediate monetary incentive. Using co-design techniques[69] for developing study apps involving researchers and participants could help guide the development of most parsimonious research protocols that fit into the daily lives of people and are still sufficiently comprehensive for researchers.

In the present diverse sample of user-activity data, several cohort characteristics such as age, disease status, clinical referral, monetary benefits, etc have emerged as key drivers for higher retention. These characteristics may also guide the development of new data-driven engagement strategies[70,71] such as tailored just-in-time interventions[72] targeting sub-populations that are most likely to drop out early from remote research. Left unchecked the ongoing bias in participant recruitment combined with inequitable long-term participation in large scale "digital cohorts" can severely impact the generalizability[36,37] and undermine the promise of digital health in collecting representational real-world data.

4.6 ACKNOWLEDGMENTS

Funding Support: This work was supported in part by various funding agencies that included the National Institute of Mental Health (MH100466), Robert Wood Johnson Foundation (RWJF - 73205), National Library of Medicine (R01LM013043), National Center for Advancing Translational Sciences (1UL1TR002319-01) and American Sleep Apnea Foundation, Washington, DC. The funding agencies did not play a role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit this article for publication. We also would like to thank Phendo[81], SleepHealth[82], and GoodRx[83] groups for sharing user engagement data from their respective digital health studies for the retention analysis.

We also acknowledge and thank researchers from previously completed digital health studies[16,23,84–86] for making the user-engagement data available to the research community under qualified researcher program[87]. Specifically, mPower study data were contributed by users of the Parkinson mPower mobile application as part of the mPower study developed by Sage Bionetworks and described in Synapse [doi:10.7303/syn4993293]. MyHeartCounts study data were contributed by users of the My Heart Counts Mobile Health Application study developed by Stanford University and described in Synapse [doi:10.7303/syn11269541]. Asthma study data were contributed by users of the asthma app as part of the Asthma Mobile Health Application study developed by The Icahn School of Medicine at Mount Sinai and described in Synapse [doi:10.7303/syn8361748]. We would like to thank and acknowledge all study participants for contributing their time and effort to participate in these studies.

4.7 REFERENCES

1. Fogel, D. B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. Contemporary Clinical Trials Communications 11, 156–164 (2018).

2. Briel, M. et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. J. Clin. Epidemiol. 80, 8–15 (2016).

3. Bradshaw, J., Saling, M., Hopwood, M., Anderson, V. & Brodtmann, A. Fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease is qualitatively distinct. J. Neurol. Neurosurg. Psychiatry 75, 382–387 (2004).

4. Snyder, M. & Zhou, W. Big data and health. The Lancet Digital Health (2019). doi:10.1016/s2589-7500(19)30109-8

5. Sim, I. Mobile Devices and Health. N. Engl. J. Med. 381, 956–968 (2019).

6. Steinhubl, S. R., Muse, E. D. & Topol, E. J. The emerging field of mobile health. Sci. Transl. Med. 7, 283rv3 (2015).

7. ElZarrad, M. K., Khair ElZarrad, M. & Corrigan□Curay, J. The US Food and Drug Administration's Real□World Evidence Framework: A Commitment for Engagement and Transparency on Real□World Evidence. Clinical Pharmacology & Therapeutics (2019). doi:10.1002/cpt.1389

8. May, M. Clinical trial costs go under the microscope. Nature Medicine (2019). doi:10.1038/d41591-019-00008-7

9. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center's Global Attitudes Project (2019). Available at: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/. (Accessed: 30th July 2019)

10. Turner, B. A. Smartphone Addiction & Cell Phone Usage Statistics in 2018. BankMyCell (2018). Available at: https://www.bankmycell.com/blog/smartphone-addiction/. (Accessed: 30th July 2019)

11. Global time spent on social media daily 2018 | Statista. Statista Available at: https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/. (Accessed: 30th July 2019)

12. Steinhubl, S. R. & Topol, E. J. Digital medicine, on its way to being just plain medicine. NPJ Digit Med 1, 20175 (2018).

13. Perry, B. et al. Use of Mobile Devices to Measure Outcomes in Clinical Research, 2010-2016: A Systematic Literature Review. Digit Biomark 2, 11–30 (2018).

14. Trister, A. D., Ray Dorsey, E. & Friend, S. H. Smartphones as new tools in the management and understanding of Parkinson's disease. npj Parkinson's Disease 2, (2016).

15. Dorsey, E. R. et al. The Use of Smartphones for Health Research. Acad. Med. 92, 157–160 (2017).

16. Bot, B. M. et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 3, 160011 (2016).

17. Webster, D. E. et al. The Mole Mapper Study, mobile phone skin imaging and melanoma risk data collected using ResearchKit. Scientific Data 4, (2017).

18. Crouthamel, M. et al. Using a ResearchKit Smartphone App to Collect Rheumatoid Arthritis Symptoms From Real-World Participants: Feasibility Study. JMIR mHealth and uHealth 6, e177 (2018).

19. McConnell, M. V. et al. Feasibility of Obtaining Measures of Lifestyle From a Smartphone App: The MyHeart Counts Cardiovascular Health Study. JAMA Cardiol 2, 67–76 (2017).

20. Chan, Y.-F. Y. et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat. Biotechnol. 35, 354–362 (2017).

21. McKillop, M., Mamykina, L. & Elhadad, N. Designing in the Dark: Eliciting Self-tracking Dimensions for Understanding Enigmatic Disease. in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18 1–15 (ACM Press, 2018).

22. Waalen, J. et al. Real world usage characteristics of a novel mobile health self-monitoring device: Results from the Scanadu Consumer Health Outcomes (SCOUT) Study. PLoS One 14, e0215468 (2019).

23. Pratap, A. et al. Using Mobile Apps to Assess and Treat Depression in Hispanic and Latino Populations: Fully Remote Randomized Clinical Trial. J. Med. Internet Res. 20, e10130 (2018).

24. Anguera, J. A., Jordan, J. T., Castaneda, D., Gazzaley, A. & Areán, P. A. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2, 14–21 (2016).

25. Smalley, E. Clinical trials go virtual, big pharma dives in. Nat. Biotechnol. 36, 561–562 (2018).

26. Virtual Clinical Trials Challenges and Opportunities: Proceedings of a Workshop : Health

and Medicine Division. Available at: http://www.nationalacademies.org/hmd/Reports/2019/virtual-clinical-trials-challenges-and-opportunities-pw.aspx. (Accessed: 12th August 2019)

27. Orri, M., Lipset, C. H., Jacobs, B. P., Costello, A. J. & Cummings, S. R. Web-based trial to evaluate the efficacy and safety of tolterodine ER 4 mg in participants with overactive bladder: REMOTE trial. Contemp. Clin. Trials 38, 190–197 (2014).

28. Moore, S. et al. Consent Processes for Mobile App Mediated Research: Systematic Review. JMIR Mhealth Uhealth 5, e126 (2017).

29. Althubaiti, A. Information bias in health research: definition, pitfalls, and adjustment methods. J. Multidiscip. Healthc. 9, 211–217 (2016).

30. Druce, K. L., Dixon, W. G. & McBeth, J. Maximizing Engagement in Mobile Health Studies: Lessons Learned and Future Directions. Rheum. Dis. Clin. North Am. 45, 159–172 (2019).

31. Christensen, H., Griffiths, K. M., Korten, A. E., Brittliffe, K. & Groves, C. A Comparison of Changes in Anxiety and Depression Symptoms of Spontaneous Users and Trial Participants of a Cognitive Behavior Therapy Website. Journal of Medical Internet Research 6, e46 (2004).

32. Christensen, H., Griffiths, K. M. & Jorm, A. F. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 328, 265 (2004).

33. Eysenbach, G. The Law of Attrition. Journal of Medical Internet Research 7, e11 (2005).

34. Christensen, H. & Mackinnon, A. The law of attrition revisited. Journal of medical Internet research 8, e20; author reply e21 (2006).

35. Eysenbach, G. The Law of Attrition Revisited – Author's Reply. Journal of Medical Internet Research 8, e21 (2006).

36. Kaplan, R. M., Chambers, D. A. & Glasgow, R. E. Big data and large sample size: a cautionary note on the potential for bias. Clin. Transl. Sci. 7, 342–346 (2014).

37. Neto, E. C. et al. A Permutation Approach to Assess Confounding in Machine Learning Applications for Digital Health. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19 (2019). doi:10.1145/3292500.3330903

38. Dainesi, S. M. & Goldbaum, M. Reasons behind the participation in biomedical research: a brief review. Rev. Bras. Epidemiol. 17, 842–851 (2014).

39. Balachandran, M. Way To Health. Available at: https://www.waytohealth.org/. (Accessed: 17th August 2019)

40. Bentley, J. P. & Thacker, P. G. The influence of risk and monetary payment on the research participation decision making process. J. Med. Ethics 30, 293–298 (2004).

41. Laport-López, F., Serrano, E., Bajo, J. & Campbell, A. T. A review of mobile sensing systems, applications, and opportunities. Knowledge and Information Systems (2019). doi:10.1007/s10115-019-01346-1

42. Laursen, D. R. T., Paludan-Müller, A. S. & Hróbjartsson, A. Randomized clinical trials with run-in periods: frequency, characteristics and reporting. Clin. Epidemiol. 11, 169–184 (2019).

43. Pablos-Méndez, A., Barr, R. G. & Shea, S. Run-in periods in randomized trials: implications for the application of results in clinical practice. JAMA 279, 222–225 (1998).

44. Dickert, N. & Grady, C. What's the Price of a Research Subject? Approaches to Payment for Research Participation. New England Journal of Medicine 341, 198–203 (1999).

45. Gelinas, L. et al. A Framework for Ethical Payment to Research Participants. N. Engl. J. Med. 378, 766–771 (2018).

46. A guide to actively involving young people in research – INVOLVE. Available at: https://www.invo.org.uk/posttypepublication/a-guide-to-actively-involving-young-people-in-research/. (Accessed: 27th August 2019)

47. Nicholson, L. M. et al. Recruitment and retention strategies in longitudinal clinical studies with low-income populations. Contemporary Clinical Trials 32, 353–362 (2011).

48. Nicholson, L. M., Schwirian, P. M. & Groner, J. A. Recruitment and retention strategies in clinical studies with low-income and minority populations: Progress from 2004–2014. Contemporary Clinical Trials 45, 34–40 (2015).

49. Druce, K. L. et al. Recruitment and Ongoing Engagement in a UK Smartphone Study Examining the Association Between Weather and Pain: Cohort Study. JMIR mHealth and uHealth 5, e168 (2017).

50. Liu, H.-E. & Li, M.-C. Factors influencing the willingness to participate in medical research: a nationwide survey in Taiwan. PeerJ 6, e4874 (2018).

51. Shavers, V. L., Lynch, C. F. & Burmeister, L. F. Factors that influence African-Americans' willingness to participate in medical research studies. Cancer 91, 233–236 (2001).

52. Trauth, J. M., Musa, D., Siminoff, L., Jewell, I. K. & Ricci, E. Public Attitudes Regarding

Willingness to Participate in Medical Research Studies. Journal of Health & Social Policy 12, 23–43 (2000).

53. 5 key findings about public trust in scientists in the U.S. Pew Research Center Available at: https://www.pewresearch.org/fact-tank/2019/08/05/5-key-findings-about-public-trust-in-scientists-in-the-u-s/. (Accessed: 11th August 2019)

54. Nebeker, C., Murray, K., Holub, C., Haughton, J. & Arredondo, E. M. Acceptance of Mobile Health in Communities Underrepresented in Biomedical Research: Barriers and Ethical Considerations for Scientists. JMIR Mhealth Uhealth 5, e87 (2017).

55. Christopher Gibbons, M. Use of health information technology among racial and ethnic underserved communities. Perspect. Health Inf. Manag. 8, 1f (2011).

56. Massoud, M. R., Rashad Massoud, M. & Kimble, L. Faculty of 1000 evaluation for Global burden of diseases, injuries, and risk factors for young people's health during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. F1000 - Post-publication peer review of the biomedical literature (2016). doi:10.3410/f.726353640.793525182

57. Oh, S. S. et al. Diversity in Clinical and Biomedical Research: A Promise Yet to Be Fulfilled. PLOS Medicine 12, e1001918 (2015).

58. Kondo, N. et al. Income inequality, mortality, and self rated health: meta-analysis of multilevel studies. BMJ 339, b4471 (2009).

59. Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H. & Subramanian, S. V. Race/Ethnicity, Gender, and Monitoring Socioeconomic Gradients in Health: A Comparison of Area-Based Socioeconomic Measures—The Public Health Disparities Geocoding Project. American Journal of Public Health 93, 1655–1671 (2003).

60. Roth, G. A. et al. Trends and Patterns of Geographic Variation in Cardiovascular Mortality Among US Counties, 1980-2014. JAMA 317, 1976–1992 (2017).

61. All of Us Research Program Investigators et al. The 'All of Us' Research Program. N. Engl. J. Med. 381, 668–676 (2019).

62. Rogers, E. M., Singhal, A. & Quinlan, M. M. Diffusion of Innovations 1. An Integrated Approach to Communication Theory and Research 415–434 (2019). doi:10.4324/9780203710753-35

63. Karni, E. & Safra, Z. 'Preference Reversals' and the Theory of Decision Making under Risk. Risk, Decision and Rationality 163–172 (1988). doi:10.1007/978-94-009-4019-2_11

64. Greenhalgh, T. et al. Introduction of shared electronic records: multi-site case study using diffusion of innovation theory. BMJ 337, a1786 (2008).

65. Ward, R. The application of technology acceptance and diffusion of innovation models in healthcare informatics. Health Policy and Technology 2, 222–228 (2013).

66. Emani, S. et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J. Med. Internet Res. 14, e150 (2012).

67. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy & Committee on the Return of Individual-Specific Research Results Generated in Research Laboratories. Returning Individual Research Results to Participants: Guidance for a New Research Paradigm. (National Academies Press (US), 2018).

68. Mobile Apps: What's A Good Retention Rate? Available at: http://info.localytics.com/blog/mobile-apps-whats-a-good-retention-rate. (Accessed: 13th August 2019)

69. Yardley, L., Morrison, L., Bradbury, K. & Muller, I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J. Med. Internet Res. 17, e30 (2015).

70. O'Connor, S. et al. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Medical Informatics and Decision Making 16, (2016).

71. Pagoto, S. & Bennett, G. G. How behavioral science can advance digital health. Transl. Behav. Med. 3, 271–276 (2013).

72. Nahum-Shani, I. et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann. Behav. Med. 52, 446–462 (2018).

73. ResearchKit and CareKit. Apple Available at: http://www.apple.com/researchkit/. (Accessed: 29th August 2019)

74. Bewick, V., Cheek, L. & Ball, J. Statistics review 12: Survival analysis. Crit. Care 8, 389 (2004).

75. Bland, J. M. & Altman, D. G. The logrank test. BMJ 328, 1073 (2004).

76. Rich, J. T. et al. A practical guide to understanding Kaplan-Meier curves. Otolaryngol. Head Neck Surg. 143, 331–336 (2010).

77. Parsons, S. Introduction to Machine Learning by Ethem Alpaydin, MIT Press, 0-262-01211-1, 400 pp., $50.00/£32.95. The Knowledge Engineering Review 20, 432–433 (2005).

78. US Census Bureau. 2018 National and State Population Estimates.

79. Kumar, D. & Klefsjö, B. Proportional hazards model: a review. Reliability Engineering & System Safety 44, 177–188 (1994).

80. R: The R Project for Statistical Computing. Available at: http://www.R-project.org/. (Accessed: 13th March 2019)

81. Phendo. Available at: http://citizenendo.org/phendo/. (Accessed: 1st October 2019)

82. SleepHealth Study and Mobile App - SleepHealth. SleepHealth Available at: https://www.sleephealth.org/sleephealthapp/. (Accessed: 1st October 2019)

83. Website. Available at: https://www.goodrx.com/. (Accessed: 1st October 2019)

84. Chan, Y.-F. Y. et al. The asthma mobile health study, smartphone data collected using ResearchKit. Sci Data 5, 180096 (2018).

85. Hershman, S. G. et al. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the MyHeart Counts Study. Sci Data 6, 24 (2019).

86. Arean, P. A. et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J. Med. Internet Res. 18, e330 (2016).

87. Wilbanks, J. & Friend, S. H. First, design for data sharing. Nature biotechnology 34, 377–379 (2016).

Chapter 5. INDIVIDUALS' WILLINGNESS TO PARTICIPATE AND SHARE DIGITAL DATA IN ONLINE BIOMEDICAL RESEARCH

5.1 ABSTRACT

Importance: Using social media to recruit participants is a common and cost-effective practice. Willingness to participate (WTP) in biomedical research is a function of trust in the scientific team, which is closely tied to the source of funding and institutional connections.

Objective: To determine if WTP and willingness to share social media data varies by the research team and online recruitment platform.

Design: Mixed methods longitudinal survey conducted over two timepoints (T1 and T2).

Setting: Conducted online using Amazon's MTurk platform.

Participants: Participants were US-dwelling adults 18 years or older who use at least one social media platform. Recruitment was stratified to match race/ethnicity proportions of the 2010 US census. The volunteer sample consisted of 914 participants at Time 1; 655 completed the follow-up survey five months later.

Main Outcomes and Measures: Outcomes were (1) past experience with online research and sharing social media data for research; (2) WTP in research advertised online; (3) WTP in a study sponsored by a pharmaceutical company, a university, or a federal agency; and (3) willingness to share social media data. We also probed opinions regarding GDPR, which came into effect between T1 and T2.

Results: The 914 participants completing the first survey (T1) were relatively young (66.1% aged 18-39 years old) and 54% female. Of these, 655 participants responded at T2. While 74.4%

indicated WTP in biomedical research, only 49.3% were willing to share their social media data. Participants were significantly more likely to participate ($P < .0001$) and share their social medial data ($P < .0001$) in university-led research compared to federally or pharma-led research. WTP in pharma-led research declined (T2-T1 = -11·89%, $P < .0001$). Reasons for WTP were interest in furthering science, financial incentives, trust in the organization, and data security. While 63% of respondents reported seeing new privacy policy emails related to the new GDPR law, only 27% indicated this positively influenced their WTP. Thematic analysis of responses indicated that WTP may improve with stronger data security measures.

Conclusion and Relevance: Researchers may see reduced online research participation and data sharing, particularly for research conducted outside academia.

5.2 INTRODUCTION

With nine-in-ten Americans seeking information on the web[1] and seven-in-ten using social media platforms[2], the use of online mediums to recruit and to collect research data from diverse populations has become a common and cost-effective practice in health sciences research over the last five years[3-6] This form of recruitment and data collection is currently in use in large scale biomedical research projects, such as the NIH's Precision Medicine Initiative[7] (AllofUS™), which plans to recruit a diverse sample of 1,000,000 Americans through social media campaigns. Such projects also intend to collect digital information (electronic health records information, data from fitness devices, and even social media and web-searches) to enhance our understanding of early risk factors for different disease states. Even social media companies are using digital data to inform better outcomes; for instance, Facebook has been able to use social media data to identify suicide risk in their users, and as a result, has formed a Compassion Team to address these issues[8].

Recent data privacy violations[9–11], potentially threaten the ability for biomedical researchers to recruit participants through online platforms and collect digital data from participants. Paramount to recruitment and subsequent participation in biomedical research is participant trust in science, the investigative team, and the management of personal information. Generations of biomedical research misconduct such as the Tuskegee Syphilis Study have influenced the public's trust in biomedical research[12]. A recent Pew Charitable Trust survey of trust in the internet found that even experts in digital security were mixed in their impressions that the general population will continue to share personal data online, with less than 50% of experts saying trust will improve with new regulations, and the remainder indicating that it will stay the same or erode over time[13]. Another study from Australia found that while patients still feel that sharing personal information is important for biomedical research, there are considerable concerns voiced about how the data will be managed and that willingness to share such data is dependent on who is collecting the data[14]. Lack of trust in studies advertised via the internet and social media and concerns about data security may bias samples collected in this manner[15]. As a result, the use of these platforms for recruitment and data collection for biomedical research raises significant data privacy, ethics, ownership and stewardship challenges[16] for IRBs, researchers, and participants.

The purpose of this mixed-methods study was to ascertain (1) the general populations' willingness to participate in biomedical research advertised on different digital platforms, (2) to determine if the study sponsor further modified the decision and willingness to participate, (3) whether people are willing to share digital data in biomedical research, and (4) whether willingness to participate improves with announcements regarding new data privacy laws[17].

5.3 METHODS

5.3.1 *Recruitment and Eligibility*

Participants were recruited using Amazon's Mechanical Turk (MTurk) platform[18]. MTurk is an online crowdsourcing platform where workers are paid to complete tasks such as data processing, problem-solving, and surveys. The platform is regularly used in health research[19] and allows investigators to sample study participants from a large representative and more diverse population[20] than typically seen in an in-person study at a fraction of the cost and time.

To be eligible, participants had to live in the U.S., be 18 years old or older, and use at least one social media platform. To ensure we were recruiting appropriate participants from the United States, we set the MTurk survey criteria to only include workers who lived and graduated high school in the United States (see eAppendix-1 for screening questions). The participant recruitment was stratified to match race/ethnicity proportions to that of the 2010 US census data[21].

5.3.2 *Procedures*

The University of Washington Institutional Review Board gave this study a category 2 exempt status because this is an opinion survey with participants the investigator cannot identify[22]. Participants were provided with a brief explanation of the survey on the MTurk platform; Participants were also informed that the team would contact them again in approximately 3 months to take a follow-up survey, which was also completely voluntary. Information about compensation was provided for T1 ($3) and T2 ($5) surveys. Once they consented, participants were asked to provide preliminary demographic information to determine eligibility. MTurk platform was used to deploy the survey developed using REDCap (Research Electronic Data Capture)[23] hosted at the Institute of Translational Health Sciences (ITHS), University of Washington. REDCap is a secure,

web-based application designed to support data capture for clinical and research studies developed through a multi-institutional collaborative effort. The first survey (T1) was administered in April 2018. The second survey (T2) was sent in September 2018 to all participants that completed the first survey. The primary goal of the T2 survey was to assess stability of WTP over time and allowed us to assess the impact of the European General Data Protection Regulation (GDPR) Law[24] on WTP, that came into effect in May 2018. Participants were given up to three reminders to complete the second survey. See Figure 5.1 for an overview of the study procedures and supplementary materials(Appendix B) for the Screening, T1 and T2 surveys.

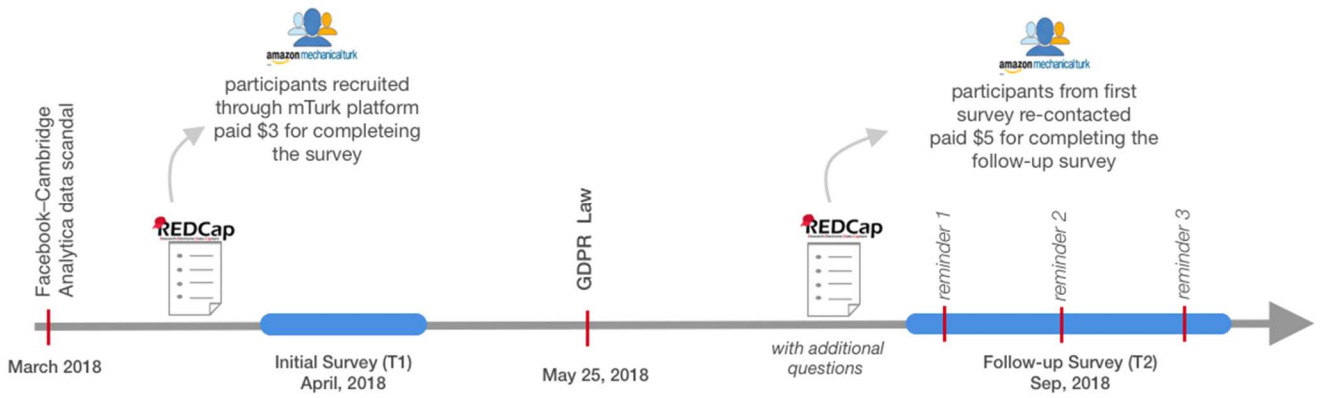


Figure 5.1. Overall schematic of the study design.

The initial survey (T1) was deployed through Amazon's MTurk platform, a month after the news about Facebook-Cambridge Analytica data privacy violations surfaced. The second survey (T2) was sent five months later in September 2018 to all the participants who responded to the T1 survey. A total of three reminders were sent to participants for completing the T2 survey.

Demographic Information: Demographic (sex, race/ethnicity, age, education) and social media use were self-reported by participants. Participants were also asked if they had ever volunteered for an online study before and if they had ever shared social media data for research purposes.

Survey Questions: The survey was developed by the authors and pilot tested to ensure clarity and understanding. The outcomes of interest were (1) participants' past experience with online research, and whether they had ever shared social media data for research purposes; (2) willingness to participate in biomedical research advertised on Google Search or Facebook; (3) willingness to participate in a study sponsored by a pharmaceutical company (e.g., Pfizer), a university (e.g., University of California, Los Angeles), or a federal agency (e.g., The National Institutes of Health); and (3) willingness to share social media data with a study sponsored by a pharmaceutical company, a university, or a federal agency. During T2, we also included questions about the new European General Data Protection Regulation law (GDPR), which came into effect on May 25, 2018, between the time T1 and T2. Outcomes of interest were (1) whether participants had noticed emails from social media companies related to the GDPR law and (2) whether this new law reassured them about data security. For each question, participants were given an opportunity to explain the reason for their answers in an open field text box. See eAppendix 2-3 for a copy of the T1 and T2 surveys.

5.3.3 *Data Analysis*

Participants' responses to structured survey questions were summarized using summary statistics. Differences in demographics between the participants that completed the first survey (T1) and a subset that responded to the second survey (T2) were assessed using a Chi-square test. We have used a conservative minimum response rate (RR_1) based on AAPOR reporting guidelines[25] to

report the participant response rate (RR) for the T2 survey. Participant responses to the main outcomes of interest (WTP and willingness to share social media data) across the two survey timepoints were evaluated using a logistic regression model based on generalized estimating equations (GEE)[26]. Briefly, GEE is a semi-parametric method to estimate population-averaged effects by accounting for correlations in time-invariant data (that is, participant responses over time T1 and T2) using robust and unbiased standard errors. We also accounted for differences in participant responses due to demographics such as age, gender, and race/ethnicity. Due to small sub-group sample sizes, race/ethnicity was collapsed into a binary variable of minority/non-minority and participants within age groups 55-69 years and above 70 were collapsed into one 55 and over age group. To assess the stability of response over time, an interaction term indicating survey time (T1 vs T2) was included for each covariate in the GEE model. We also assessed the combined effect of the recruitment platform and study sponsors on WTP and data sharing using an interaction term. The significance (*P* values) of the model estimates were corrected for multiple testing using the FDR method.

A mixed-methods approach combined quantitative and qualitative data with the function of expansion[27], allowing inductive qualitative data to provide the "why" to questions uncovered by the quantitative data. Missing data were not included in the analysis. Qualitative data were imported into Dedoose[28] and analyzed using thematic analysis[29]. The survey was developed based on recent news events of social media data breaches and mishandling, with a pragmatic interest in how such public discourse may influence participant recruitment and retention for studies. Two coders independently familiarized themselves with the data and then coded a portion of survey responses to extract initial themes. Themes were developed and revised until saturation was

reached. The themes were independently arrived at by the first two coders, and then verified by additional coders. Data were iteratively reviewed (open coded) and collapsed to mutually exclusive themes (axial coding). For the second survey, we confirmed T1 themes, while still allowing for new themes to emerge. Triangulation[30] of quantitative and qualitative data allowed for convergence of themes and a more comprehensive understanding of WTP and willingness to share social media data. Illustrative quotes and themes are provided for a qualitative data audit trail. No power analysis was conducted, as this exploratory study did not attempt to demonstrate the effects of a particular magnitude and no similar standards of sample size exist for qualitative studies. Rather, we collected a sample large enough to contribute new knowledge to the analysis; during coding, saturation was achieved when no new themes emerged[31]. All quantitative analysis was done using R[32] statistical programming language.

5.4 RESULTS

Table 5.1. Comparison of participant demographics across the two surveys conducted in April (T1) and September (T2) 2018

	Survey Time Period		
	T1: April 2018	**T2: September 2018**	***P*-value**
N	914	655	
Age (%)			0.97
18-24	76 (8.3)	51 (7.8)	
25-39	528 (57.8)	379 (57.9)	
40-54	226 (24.7)	167 (25.5)	
55-69	72 (7.9)	48 (7.3)	
70+	12 (1.3)	10 (1.5)	
Gender - Female (%)	494 (54.0)	346 (52.8.)	0.67
Race/Ethnicity (%)			0.84
White	615 (67.3)	439 (67.0)	
Hispanic/Latino	127 (13.9)	82 (12.5)	
Black/African American	107 (11.7)	86 (13.1)	
Asian	52 (5.7)	40 (6.1)	
Hawaiian/Pacific Islander/Native American/Alaska Native	13 (1.4)	8 (1.2)	

5.4.1 *Sample Characteristics*

A total of 985 participants were recruited at time 1 (T1). Of these, 655 (66.5% RR_1) participants responded to the time 2 (T2) survey. Responses from 71 (7.2%) participants were excluded from the data analysis due to questionable data (e.g., duplicate responses across questions, pasting of irrelevant text). No significant differences were seen in the participant demographics across the two surveys (Table 5.1). Overall, the cohort was relatively young, with 66.1% aged 18-39. The majority (67.3%) reported being Non-Hispanic White followed by Hispanic/Latino (13.9%) and

African American (11.7%); approximately half (54%) were female. The majority (72%) indicated that they had participated in online research previously, with 23% of that subsample stating they had shared social media data for research purposes.

5.4.2 *Time 1 Analyses*

Willingness to participate (WTP) in biomedical research by recruitment platform and sponsor. We identified significant differences in willingness to participate in research by recruitment platform and by the study sponsor. 74.4% (*n* =680) of the sample indicated WTP in a biomedical research study run by one of the three institutions (either a university, a federal agency, or a pharmaceutical company). Participants were almost twice as likely to report WTP in a study sponsored by a university than they were in a study sponsored by a federal agency (OR = 0.58, 95% confidence interval [CI] = 0.51-0.64, $P < .0001$) or a pharmaceutical company (OR = 0.59, 95% CI = 0.53-0.66, $P < .0001$). The WTP was also significantly lower for older participants (those aged 55 years and older; OR = 0.36, 95% CI = 0.22-0.61, $P < .0001$) compared to young adults aged 18-24 years old. WTP was also significantly higher (OR = 1.24, 95% CI = 1.1-1.41, $P < .001$) for recruitment through Google compared to Facebook ads (university-sponsored: 61.6% vs 56.5%, federal agency-led: 49.5% vs. 43.5%, and pharmaceutical-led: 47.9% vs. 42.9%, respectively). No significant differences in WTP were observed by participant gender or race/ethnicity (Table 5.2).

Common themes derived from our qualitative analysis found that respondents were willing to participate (1) for altruistic reasons, (2) financial incentives, and (3) trust/credibility of the sponsor. Themes regarding disincentive to participate were concerns about data security and lack of trust

in the study sponsor. See eAppendix-6 for illustrative participant quotes that represent these themes.

Willingness to share social media data. Fifty one percent of the sample (n = 464) preferred not to share their social media data with any entity. The remaining participants (49.3%; n = 454) were willing to share their data with at least one of the three study sponsors. Of those willing to share, 23.9% (n = 219) were willing to share with all three, 13.1% (n = 120) with two of the three sponsors, and the rest (12.1%; n = 111) with only one institution. Participants were significantly more likely to share their social media data in university-led research (45.0% of the respondents) compared to research sponsored by a federal agency (35.2% of the respondents; OR = 0.65, 95% CI = 0.58-0.72, $P < 0.0001$) or pharmaceutical-sponsored research (29.5% of the respondents; OR = 0.50, 95% CI = 0.44-0.56, $P < 0.0001$). Willingness to share social media data was also lower for middle-aged (40-54 years; OR = 0.46, 95% CI = 0.28-0.74, $P < .001$) and older adult participants (55 years and older; OR = 0.37, 95% CI = 0.20-0.69, $P < .001$) compared to younger adults aged 18-39 years old. No significant difference in willingness to share by race/ethnicity or gender was observed (Table 3). Major themes here were similar to themes for WTP, with universities being seen as trustworthy, but the trustworthiness of pharmaceutical and federal sponsors was questioned. See eAppendix-7 for illustrative participant quotes that represent these themes.

5.4.3 *Time 2 Analysis*

Willingness to participate (WTP) 655 (66.5% RR_1) participants responded to the T2 survey. WTP only changed for pharmaceutical-sponsored research, which decreased by T2 (T2-T1 = -11.89%, OR = 0.62, 95% CI = 0.54-0.77, $P < .0001$; see Table 5.2-a). Older participants (55 years

and older) who responded at T2 showed significantly higher willingness to participate (increase in OR: 2.95, 95% CI=1.5-5.9, $P < .001$) compared to their WTP during T1 (Table 5.2). Participant preference for recruitment via Google ads as observed in T1 (OR = 1.24) also reduced over time (decrease in OR = 0.77, 95% CI=0.5-0.9, $P < .05$) and was nearly the same as the Facebook platform ($OR_{Google\text{-}vs\text{-}Facebook(T2)}$:0.96) (See eAppendix 4-5 for further breakdown of group-wise proportions). No new themes emerged between T2 and T1 regarding WTP.

Willingness to share social media data Willingness to share social media data declined significantly for all but university-led studies. While 43.1% of T2 respondents were willing to share social media with university-led studies, willingness to share with pharmaceutical companies dropped to 29.5% (-6.84%, $P = .01$) and 35.3% with federally-led research studies (-7.32%, $P = .01$ respectively; Table 5.2-b). Continued privacy and data security concerns reported in the news were noted as a problem in the qualitative data.

Table 5.2. Odds ratios for willingness to participate in online biomedical research at time T1 and change over time T2.

OR were determined using logistic regression based on the method of generalized estimating equations including assessing the impact of participants' demographics, study sponsor, and recruitment platform on WTP. [FDR corrected *P*-values * < .05, ** <0.001, *** < 0.0001]

	T1 Survey		**Change over time (interaction effect - T2 Survey)**	
	odds ratio	**95% CI**	**odds ratio**	**95% CI**
Intercept	2.03***	1.4-2.95		
Survey-T2	0.78	0.45-1.35		
Sponsor				
University	1.0			
Pharma	0.58***	0.51-0.64	0.62***	0.54-0.77
Federal	0.59***	0.53-0.66	0.84	0.71-1.0
Platform				
Facebook	1.0			
Google	1.24**	1.1-1.41	0.77**	0.64-0.92
Age				
18-24	1.0			
25-39	0.62*	0.43-0.89	1.47	0.86-2.49
40-54	0.54*	0.36-0.81	1.93	1.09-3.41
55 and over	0.36***	0.22-0.61	2.95**	1.46-5.92
Gender				
Female	1.0			
Male	0.97	0.79-1.2	0.94	0.73-1.22
Race				
Racial/ethnic minority	1.0			
White	1.14	0.91-1.42	0.88	0.66-1.16
Sponsor:Platform				
Pharma:Google	0.99	0.88-1.12	0.93	0.77-1.13
Federal:Google	1.03	0.91-1.15	1.11	0.94-1.32

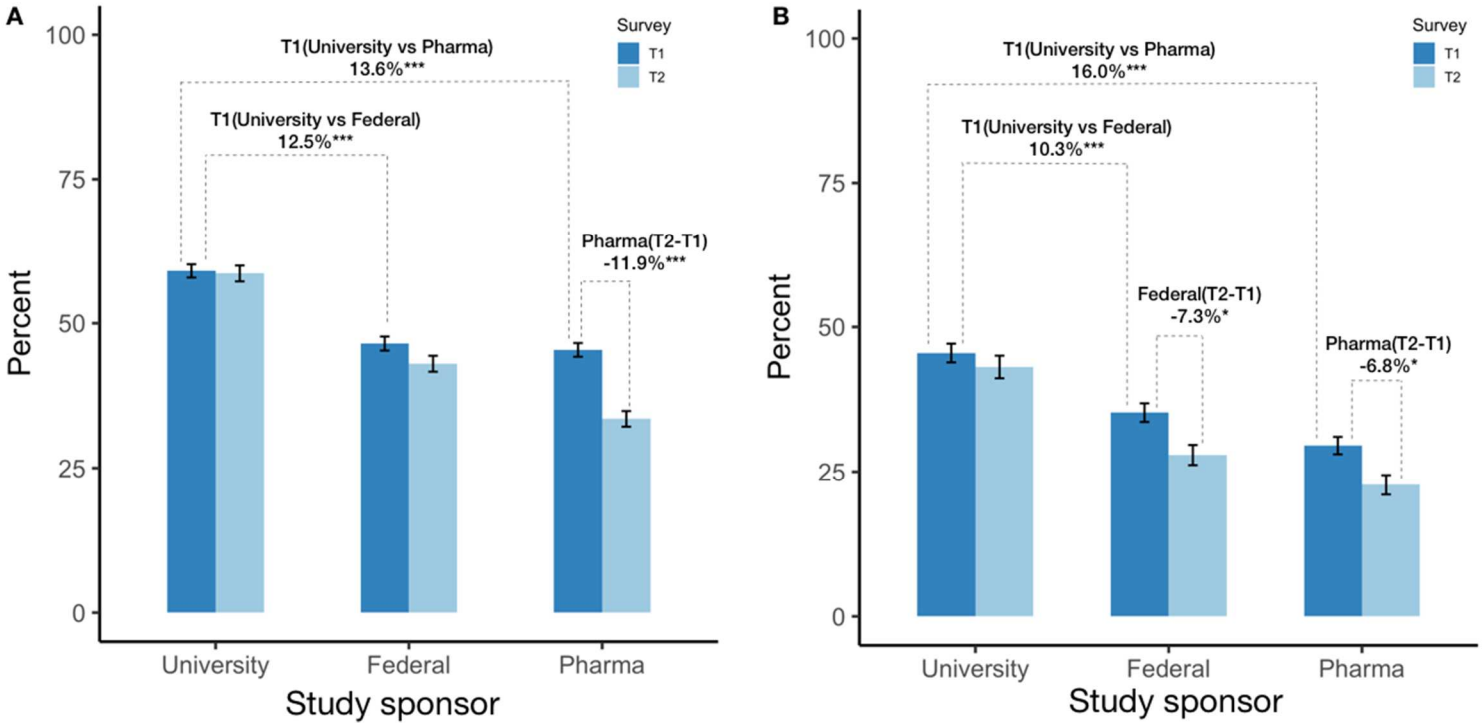


Figure 5.2. Comparing the proportion of participants willing to participate and share their social media data.

Proportions of participants faceted by the research team (university, federal agency, or pharmaceutical company) that are ***a)*** willing to participate in biomedical research and ***b)*** Share their social media data for biomedical research over time T1 and T2. Error bars indicate bootstrapped estimates of variations in participants' responses (one standard deviation). [FDR corrected *P*-values * < .05, ** <0.001, *** < 0.0001]

Impact of European GDPR Law. Sixty-three percent of participants reported seeing GDPR-related emails and/or advertisements by T2. No significant difference in WTP or willingness to share social media data was found between participants who reported seeing the GDPR-related emails and those who did not. Twenty-seven percent said that the GDPR-related messages made them feel more secure about their data and provided proof that the organization was working on

its data security. As one respondent explained, *"It shows me that they make notice of our concerns and are fixing them."* However, seventy-three percent felt the GDPR-related messages did not regain their trust. As one respondent stated: *"I think the ads are just aimed at fixing a public relations problem. They still make their money from collecting our data and selling it and they aren't going to stop."*

Table 5.3. Odds ratios for willingness to share social media data in online biomedical research at time T1 and change over time T2.

Odds ratios were determined using logistic regression based on the method of generalized estimating equations including assessing the impact of participants demographics and study sponsor on willingness to share social media data. [FDR corrected *P*-values * < .05, ** <0.001, *** < 0.0001]

	T1 Survey		**Change over time (interaction effect - T2 Survey)**	
	odds ratio	**95% CI**	**odds ratio**	**95% CI**
Intercept	1.59	1.03-2.47		
survey-T2	0.81	0.47-1.39		
Sponsor				
University	1.0			
Federal	0.65***	0.58-0.72	0.79*	0.67-0.93
Pharma	0.50***	0.44-0.56	0.76*	0.65-0.93
Age group				
18-24	1.0			
25-39	0.61	0.4-0.94	1.19	0.67-2.1
40-54	0.46**	0.28-0.74	1.58	0.86-2.92
55+	0.37**	0.2-0.69	1.58	0.71-3.52
Gender				
Female	1.00			
Male	0.98	0.77-1.25	0.82	0.6-1.12
Race				
Racial/ethnic minority	1.00			
White	0.91	0.7-1.18	0.94	0.67-1.32

5.5 DISCUSSION

Public trust in the use of digital platforms, such as Google Search and Facebook, appears to influence participants' willingness to participate in and share social media data with biomedical

research efforts. Moreover, trust in research entities is low, with a majority of participants indicating an unwillingness to share social media data with federally sponsored or pharmaceutical company-led research. Although participants acknowledged the importance of participating in biomedical research and indicated they would do so for altruistic reasons, concerns about privacy and misuse of their personal data appear to outweigh the perceived importance of volunteering to participate in such research[33]. Issues of data security and mistrust may adversely impact research projects that plan to rely on large-scale recruitment through digital platforms. Recruitment of this nature, without a concerted effort to address participant mistrust of how their data will be managed, may result in the recruitment of large yet biased samples that are not representative of the intended population. This could have a significant impact on the generalizability of outcomes from biomedical research.

Although our thematic analysis indicated that better data security measures and transparency of data use may mitigate concerns regarding participation, less than a quarter of our sample indicated that they were reassured by recent attempts at regulation such as the GDPR policies. The findings from this study are understandable in light of growing evidence that data privacy policies available on digital platforms do not accurately disclose how that information is utilized. One recent paper found that many health apps share digital data with companies like Facebook and Google, but fail to disclose this in their data privacy policies[34]. A qualitative study of participants' willingness to share research data reported similar findings, with trust in the research team and fears related to misuse arising as major concerns by potential participants[33]. Our findings, combined with others, suggest that social media campaigns and policies to address how privacy and data security will be improved may not be sufficient to address willingness to participate in online research and share digital data. As our survey results showed, participants remained untrusting of these platforms

several months after the platforms had sent out messages addressing their data security problems. However, partnership with universities and other trusted entities to develop better policies may be a useful solution, given how consistently our participants expressed trust in university-led research. As a number of studies indicate, participants' trust in research is closely linked to the institution conducting the research[35].

5.5.1 *Limitations*

The findings from this survey should be viewed with the following limitations in mind and would benefit from further research. First, this is a general population survey of participant impressions about willingness to participate in research and share personal data. To confirm our findings, a study specifically comparing recruitment avenues would need to be conducted. Second, participants were identified through MTurk, and therefore the representativeness of the findings may be influenced by our sample selection method, even though participants were recruited to match the race/ethnicity of the 2010 US census data[21]. Although MTurk participants are likely to be more aware of data sharing policies and more comfortable with online research than the general public, recent studies suggest that for research of this nature, these samples tend to be as good as, or better than in-person surveys[36]. Third, our sample was not recruited specifically to test hypotheses about racial/ethnic, gender or age differences. Particularly in regards to our age findings, that WTP seemed to improve in older populations over time, we believe this finding to be an artifact of the small sample of older adults who participated in this survey, and in all likelihood these are not participants who are representative of older adults in the general population. Thus, to truly understand demographic differences in WTP, a large study that oversamples participants from different demographic groups will need to be conducted. Finally,

the phrasing of survey questions listing depression as a health condition could also have negatively impacted study participants' WTP, given the stigma associated with the disorder[37]. Despite these limitations, however, the data are still useful for both informing recruitment practices and providing information about the concerns people have regarding the secure management of social media data for research purposes, particularly at this time.

In conclusion, willingness to participate in biomedical research advertised on social media platforms and search engines, as well as the willingness to share digital data with researchers, has been affected by recent news on the misuse of such data. Although university-led research is seen as more trustworthy than federally- or pharmaceutical company-led research, willingness to participate is still gravely impacted. Despite these concerns, social media provides opportunities for conducting biomedical research at scale[38] including enrolling minority populations[5] and help improve diversity in clinical trials, the majority of which are discontinued early due to recruitment challenges[39]. It will be important for researchers and research organizations to work more closely with participant communities to address concerns about data sharing and privacy.

5.6 ACKNOWLEDGEMENTS

Funding Support: This work was supported in part by the National Institute of Mental Health (grant numbers P50 MH115837, R01 MH102304, and R33 MH110509). The study sponsor did not play a role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit this article for publication.

5.7 REFERENCES

1. Demographics of Internet and Home Broadband Usage in the United States. *Pew Research Center: Internet, Science & Tech* (2018). Available at: https://www.pewinternet.org/fact-sheet/internet-broadband/. (Accessed: 3rd June 2019)

2. Demographics of Social Media Users and Adoption in the United States. *Pew Research Center: Internet, Science & Tech* (2018). Available at: https://www.pewinternet.org/fact-sheet/social-media/. (Accessed: 3rd June 2019)

3. Topolovec-Vranic, J. & Natarajan, K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. *J. Med. Internet Res.* **18**, e286 (2016).

4. Arean, P. A. *et al.* The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. *J. Med. Internet Res.* **18**, e330 (2016).

5. Pratap, A. *et al.* Using Mobile Apps to Assess and Treat Depression in Hispanic and Latino Populations: Fully Remote Randomized Clinical Trial. *J. Med. Internet Res.* **20**, e10130 (2018).

6. Bunge, E. L. *et al.* Facebook for recruiting Spanish- and English-speaking smokers. *Internet Interv* **17**, 100238 (2019).

7. Program Overview - All of Us | National Institutes of Health. Available at: https://allofus.nih.gov/about/about-all-us-research-program. (Accessed: 3rd June 2019)

8. Anderle, M. Making a More Empathetic Facebook. *The Atlantic* (2016). Available at: https://www.theatlantic.com/technology/archive/2016/03/facebooks-anti-bullying-efforts/473871/. (Accessed: 17th June 2019)

9. US CMS says 75,000 individuals' files accessed in data breach. *Deccan Chronicle* (2018). Available at: https://www.deccanchronicle.com/technology/in-other-news/201018/us-cms-says-75000-individuals-files-accessed-in-data-breach.html. (Accessed: 16th September 2019)

10. Perez, S. & Whittaker, Z. Everything you need to know about Facebook's data breach affecting 50M users. *TechCrunch* (2018). Available at: http://social.techcrunch.com/2018/09/28/everything-you-need-to-know-about-facebooks-data-breach-affecting-50m-users/. (Accessed: 16th September 2019)

11. Wakabayashi, D. Google Plus Will Be Shut Down After User Information Was Exposed. (2018). Available at: https://www.nytimes.com/2018/10/08/technology/google-plus-security-disclosure.html. (Accessed: 16th September 2019)

12. Kerasidou, A. Trust me, I'm a researcher!: The role of trust in biomedical research. *Medicine, Health Care and Philosophy* **20**, 43–50 (2017).

13. The Fate of Online Trust in the Next Decade. *Pew Research Center: Internet, Science & Tech* (2017). Available at: https://www.pewinternet.org/2017/08/10/the-fate-of-online-trust-in-the-next-decade/. (Accessed: 17th June 2019)

14. Krahe, M., Milligan, E. & Reilly, S. Personal health information in research: Perceived risk, trustworthiness and opinions from patients attending a tertiary healthcare facility. *J. Biomed. Inform.* 103222 (2019).

15. Guillemin, M. *et al.* Do Research Participants Trust Researchers or Their Institution? *Journal of Empirical Research on Human Research Ethics* **13**, 285–294 (2018).

16. Bender, J. L., Cyr, A. B., Arbuckle, L. & Ferris, L. E. Ethics and Privacy Implications of Using the Internet and Social Media to Recruit Participants for Health Research: A Privacy-by-Design Framework for Online Recruitment. *J. Med. Internet Res.* **19**, e104 (2017).

17. EUGDPR – Information Portal. Available at: https://eugdpr.org/. (Accessed: 4th June 2019)

18. Amazon Mechanical Turk. Available at: https://www.mturk.com/. (Accessed: 6th May 2019)

19. Créquit, P., Mansouri, G., Benchoufi, M., Vivot, A. & Ravaud, P. Mapping of Crowdsourcing in Health: Systematic Review. *J. Med. Internet Res.* **20**, e187 (2018).

20. Smith, N. A., Sabat, I. E., Martinez, L. R., Weaver, K. & Xu, S. A Convenient Solution: Using MTurk To Sample From Hard-To-Reach Populations. *Industrial and Organizational Psychology* **8**, 220–228 (2015).

21. US Census Bureau. Decennial Census Datasets.

22. Exempt Categories. *WPI* Available at: https://www.wpi.edu/research/resources/compliance/institutional-review-board/exempt-categories. (Accessed: 23rd September 2019)

23. Harris, P. A. *et al.* Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* **42**, 377–381 (2009).

24. EUGDPR – Information Portal. Available at: https://eugdpr.org/. (Accessed: 13th September 2019)

25. Standard Definitions - AAPOR. Available at: https://www.aapor.org/Publications-Media/AAPOR-Journals/Standard-Definitions.aspx. (Accessed: 29th July 2019)

26. Liang, K.-Y. & Zeger, S. L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **73**, 13 (1986).

27. Palinkas, L. A. *et al.* Mixed method designs in implementation research. *Adm. Policy Ment. Health* **38**, 44–53 (2011).

28. Home | Dedoose. Available at: https://www.dedoose.com/. (Accessed: 3rd May 2019)

29. Braun, V. & Clarke, V. Using thematic analysis in psychology. *Qualitative Research in Psychology* **3**, 77–101 (2006).

30. Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J. & Neville, A. J. The Use of Triangulation in Qualitative Research. *Oncology Nursing Forum* **41**, 545–547 (2014).

31. Malterud, K., Siersma, V. D. & Guassora, A. D. Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qual. Health Res.* **26**, 1753–1760 (2016).

32. R: The R Project for Statistical Computing. Available at: http://www.R-project.org/. (Accessed: 13th March 2019)

33. Howe, N., Giles, E., Newbury-Birch, D. & McColl, E. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *Journal of Health Services Research & Policy* **23**, 123–133 (2018).

34. Huckvale, K., Torous, J. & Larsen, M. E. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. *JAMA Netw Open* **2**, e192542 (2019).

35. Mello, M. M., Lieou, V. & Goodman, S. N. Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing. *New England Journal of Medicine* **378**, 2202–2211 (2018).

36. Golomb, B. A. *et al.* The older the better: are elderly study participants more non-representative? A cross-sectional analysis of clinical trial and observational study samples. *BMJ Open* **2**, e000833 (2012).

37. Stigma and Living with Depression | NAMI: National Alliance on Mental Illness. Available at: https://www.nami.org/Personal-Stories/Stigma-and-Living-with-Depression#. (Accessed: 24th September 2019)

38. Anguera, J. A., Jordan, J. T., Castaneda, D., Gazzaley, A. & Areán, P. A. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. *BMJ Innov* **2**, 14–21 (2016).

39. Briel, M. *et al.* A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *J. Clin. Epidemiol.* **80**, 8–15 (2016).

Chapter 6. CONCLUSIONS

The findings from my research are contemporary and extend the on-going efforts to objectively evaluate the potential fit of technology in psychiatry to help engage the general population to monitor their mental health in the real world outside the clinic. Based on the analysis of large-scale real-world datasets, my work provides additional empirical evidence that technology can help address present-day challenges in mental health care by improving access and availability of MH services and provide a new cost-effective and non-stigmatic way to assess momentary MH symptoms at population and personalized level.

Specifically, findings from *Aim-1 (Chapter 2)* further builds on "digital phenotyping" research in mental health by studying the association of smartphone usage (including informatically derived geospatial context and weather) with daily mood. My research shows that the data collected in a nationally representative and diverse sample shows significant inter-/intra-person heterogeneity between digitally determined passive features from smartphone usage and participant's self-reported mood. Additionally, the results also show empirical evidence that the participant demographics alone continue to provide a higher level of evidence predicting daily mood compared to passive data at the cohort level. While the present analysis in diverse but small cohort shows the promise of technology to assess variations in individuals' lived experience of diseases by capturing ecological real-world data over time, the results also surface the need for large sufficiently powered digital phenotyping studies (N > 10,000) to help determine robust digital biomarkers of behavioral health from the collected real-world data. However, personalized models assessing each individual's baseline show the promise of "Personalized Precision Digital

Psychiatry" by assessing an individual's drifts from their own average "digital behavior" to be a more reliable predictor of behavioral anomalies compared to a cohort level model. Finally, I also show clusters of sub-populations for whom digital traces can be indicative of behavioral fluctuations and non-responders for whom digital technology might not be the best fit for monitoring behavioral fluctuations.

Aim 2 (Chapter 3) further investigates the feasibility of deploying technology in minorities specifically Hispanic/Latinos for assessment of depression in a fully remote way study. This study was one of the first large scale effort to successfully recruit one of the largest samples of depressed Hispanic/Latinos in the United States fully remotely. While the data shows the feasibility of recruiting populations from both high and low resource settings alike by extending self-guided solutions, giving citizens back the agency and anonymity to seek help faster and at lower costs. However, the participant engagement analysis showed minorities and marginalized groups such as Hispanics/Latinos dropping our significantly earlier than their non-Hispanics/Latino counterparts, highlighting significant challenges in keeping minority populations engaged in remote online research. The findings show an urgent need for more in-depth research to understand and evaluate various participant engagement and retention strategies ("science of user attrition") to enable researchers to evaluate current and develop new methods to help enroll diverse populations and keep them engaged in remote online research.

Aim 3 (Chapter 4) investigated the specific questions about user engagement that surfaced as a result of findings from *Aim 2 (Chapter 3).* Here I empirically evaluated the flip side of real-world data collection by digital health technology i.e "*if we build the technology, will participants come*

and use it and for how long". The findings from this research surfaced key challenges in large-scale "fully remote" digital studies. Most studies were not able to recruit a nationally representative and diverse population and more than 50% of the participants left the studies within the first week. While concerning, in-depth analysis of data showed sub-groups of participants that engaged in digital studies significantly more than others highlighting the early potential fit of digital health technology for a sub-population. Also, the high user attrition seen in online research may also reflect a unique selection bias in health research in a way that is not captured by traditional in-clinic research. Combined together the findings from large-scale cross-study user engagement data could help inform the recruitment and retention strategies of future digital health research.

Finally, in *Aim 4 (Chapter 5)*, I evaluated the impact of recent data privacy and security violations on the general population's willingness to participate and share their digital data in biomedical research. The findings show growing concerns among the general public about data security and privacy violations which can negatively influence their decision to participate and share data in online remote research significantly impacting the promise of digital health. People's willingness to participate and share their digital data is also affected significantly based on the affiliation of researchers conducting the study. Additionally, people's opinions about data privacy and security did not improve as a result of new privacy laws. However, despite the concerns, the findings also show future online research could benefit by developing partnerships with universities and other trusted entities and following transparent data security and use policies.

These findings reinforce the promise of digital health technology to move the psychiatric research from subjective to objective assessment, episodic to continuous monitoring, provider-based to

ubiquitous and reactive to proactive care by offering a new approach to pragmatically assess and evaluate mental health in the real-world. However, accomplishing these goals does come with some measurable challenges[1–4] including the usability of technology by a diverse and representative population. The success of deploying digital tools in the real-world therefore will depend on both the technology's ability to detect mental health symptoms early and more importantly its social affordance and acceptance by citizens, particularly the at-risk, marginalized and ethnic minorities. With the growing minority population in the US, further research is needed to better understand socio-technical factors such as age, race/ethnicity and willingness to accept and use digital tools for remote health monitoring in a diverse population. Also, the digital characterizations of people data can he highly heterogeneous to compare at population level especially if the study sample is not large. The variability may be due to several factors such as differences in biological systems and their interactions among multiple causes, including genetic, environmental, and social, all translated through neural and developmental processes. Given the high dimensional nature of external factors that can directly or indirectly impact the behavioral health, individualized N-of-1 precision psychiatry can be a potential alternative until large scale (N ~ 10,000-100,000) studies are able to show evidence of robust digital biomarkers of behavioral health at the population level.

6.1 Future Work

Several challenges related to the technical and real-life usability of remote monitoring technology needs to be further researched in order for psychiatric researchers and clinicians to adopt digital technology to transition towards "measurement-based care" using validated digital end-points. However, data gathered remotely while large("big data") can also include several potential biases

that if left unaddressed can impact the generalizability of drawn inference significantly[5] thereby affecting our ability to develop validated digital endpoints. Future research could explore one or more of the on-going areas that at high level is tied to the 'how" to efficiently deploy and utilize the technology-based MH services in and for "who" the remote monitoring is a better fit along with "how much" and "what kinds " of RWD should be collected and for "how long".

a.) Presence of confounding in RWD: Longitudinal digital data feeds can be highly individualistic, and as a result, algorithms can learn and predict individual data structure and not necessarily the association with disease. Some of my recent joint work[6] with colleagues has shown evidence of multiple confounding factors in the RWD. We developed novel statistical methods that can quantify the "disease vs structure" learning the ability of machine learning algorithms and show that using a "person-wise" data split could help alleviate the confounding issue to an extent. Additionally, as also shown by findings from Aim-3, non-uniform participant recruitment and retention in remote digital health studies can also lead to the potential selection and ascertainment bias in the collected RWD data. Machine learning methods utilizing such data should both estimate and correct for any such confounding. As part of my joint on-going research with colleagues, we have developed a new method[7] to detect and control for confounding. The permutation-based test detects and quantifies the influence of observed confounders (demographics, case-control status, etc) and estimates the unconfounded performance of a machine learning algorithm. We evaluated the statistical properties of this method in a simulated study and real-life data from a Parkinson's disease mobile health study collected in an uncontrolled environment. New research studies could use and extend this method to evaluate the presence of confounding in the collected RWD.

b.) Data integration - Most present-day digital health studies have been underpowered to be able to detect a meaningful change given the small sample sizes compared to the high multi-dimensional nature of the RWD data. To be able to replicate and assess the variations across studies aimed at evaluating similar neuropsychiatric domains there is an urgent need to be able to integrate and compare data across studies. New data and meta-data standards that enable data sharing and re-use are needed to help the digital psychiatry move beyond early pilots to transforming clinical care.

c.) Data Security and Privacy - Given the highly personalized features of RWD combined with potential self-reported mental health assessments, its, security and privacy should be addressed with top-most priority. All studies collecting such data should clearly inform the end-user about potential risks and clearly state data storage, use, and sharing at the 6th grade reading level. Researchers could also adopt new approaches in informed-consent (e-Consent[8]) that have been utilized by large scale programs such as All of US[9] to enroll upto 1 million Americans that also includes a considerable portion of the cohort completely being enrolled and engaged "fully remotely".

d.) Technical noise in data: Despite the progress in remote data collection using smartphones, significant discrepancies still remain. For example, there are significant differences between what data types iOS and Android platforms allow the researchers to collect. These differences across platforms directly impact the data featurization, granularity, and sampling rate. iOS, for example, restricts phone and messaging logs acquisition. Also, we are still in the early stages of discovering the utility of smartphone-based passive data streams. Traditional voice calls and messaging account for < 15% of entire smartphone usage and therefore may not be the suitable proxy features

for social interaction. Basic phone usage statistics like #notifications accepted, screen usage, #app flips, keyboard strokes, etc. will help enhance the passive data feature space. However, the richness of passive data gathering comes at the cost of battery life. More research is needed to determine "how much" data is needed and at "what" sampling rate that is sufficient to allow researchers to discern patterns and at the same time balance the battery needs of passive sensing within the range acceptable to the end-users.

e.) Usability of technology: The success of technology-aided remote monitoring at its core is critically dependent on its adoption and usage by end-users. This includes study-participants and researchers in the clinical research setting and patients and providers in the delivery of healthcare. As one of the first cross-study user retention analysis conducted as part of my research shows significant challenges in participant retention, further research is needed to develop and test successful (including targeted) recruitment and retention strategies. The use of co-design methods[10] should be employed when creating digital health studies. This includes involving the intended end user in the development and conducting as is workflow analysis to ensure that the app is useful and usable and that it fits into the fabric of the person's life, not producing unnecessary burden to the end-user[11]. The digital health research could also benefit from leveraging proven theories from behavioral change research[12,13] to help improve people's adherence to remote study protocols.

f.) Utilizing new sources of real-world data: The growth and penetration of technology and smart devices continue to expand the repertoire of real-world data[14–16] that is being produced 24x7 with ever-increasing higher frequency and density (

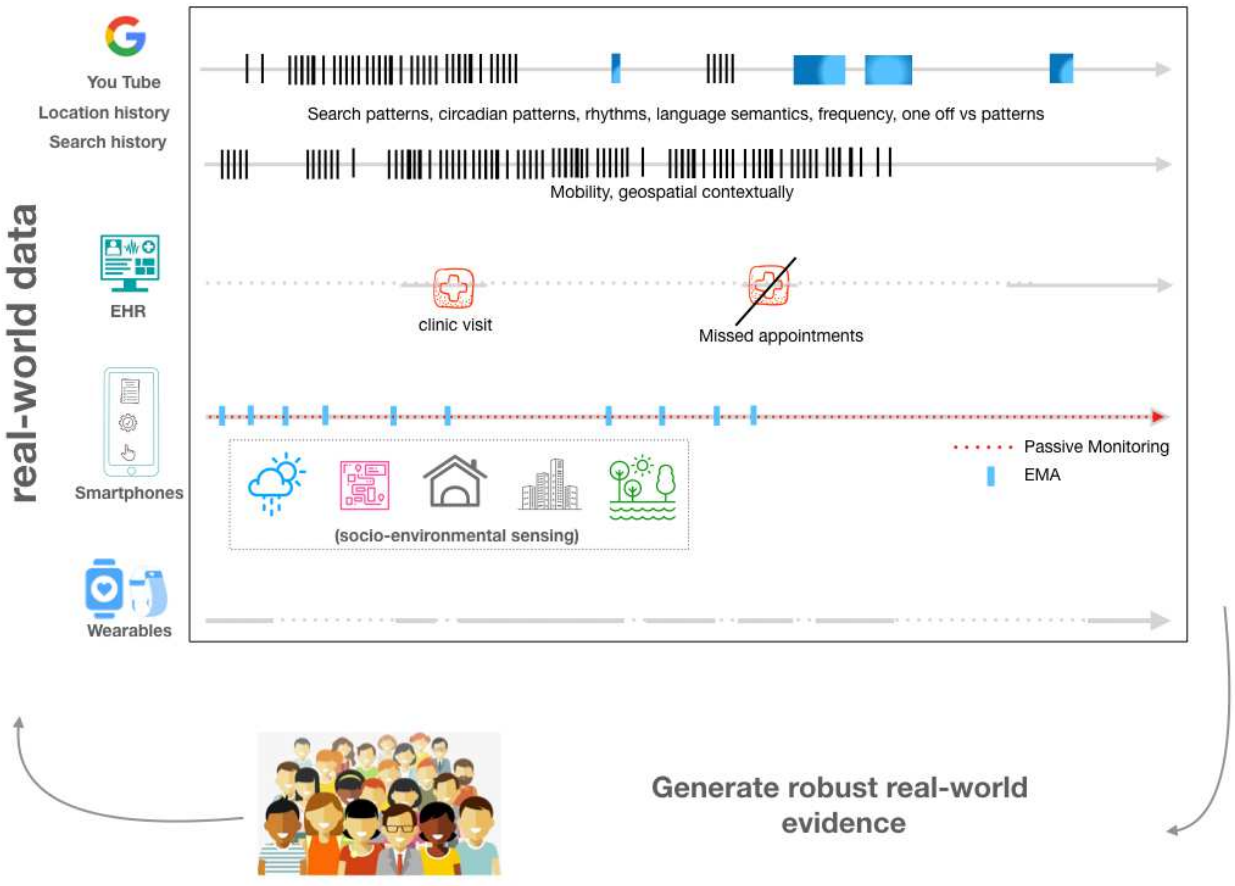

real-world data
You Tube
Location history
Search history
Search patterns, circadian patterns, rhythms, language semantics, frequency, one off vs patterns
Mobility, geospatial contextually
EHR
clinic visit
Missed appointments
Smartphones
Passive Monitoring
EMA
(socio-environmental sensing)
Wearables
Generate robust real-world evidence

Figure 6.1. Multiple sources of real-world data (RWD) for enabling future pragmatic clinical trials.). This noisy but information-packed longitudinal data sources, when triangulated at the population level, can help surface new real-world evidence that can help with early detection of various diseases including mental health disorders. While it will be out of scope to discuss the potential use cases of various new RWD sources, I here summarize two mental health-related recent use-cases. While these are early pilots, they have shown the promise to bend the curve in large scale monitoring of depression (Case A) and assess novel risk factors to help with early prediction suicidal ideation/attempt (Case B) and could be further explored in future research studies.

Case A: Using technology to offer simple "DIY" tools to aid in the self-management of mental health disorders: People are also increasingly turning towards the internet for health advice[17,18] including the communities living in the semi-urban and rural areas[19]. While there are mixed opinions[20] on the efficacy of health information received; the web as a digital medium provides a fitting opportunity for digital health researchers to deploy clinically validated assessments of MHD that are accessible through both conventional web interfaces and smartphones. The ease of using digital assessments without requiring a sign-up, joining a study, or downloading an app is a powerful *"do-it-yourself" (DIY) way,* for communities to track their mental health symptoms anonymously and receive actionable feedback; without revealing their identities and therefore avoid the stigma issues. This DIY model also empowers researchers to rapidly develop very large cross-sectional studies to examine the incidence and burden of mental illness at the population scale. Furthermore, for the subset of users that opt to share and provide more specific information such as their city/zip code data could enable a systematic assessment of temporal environmental

factors that are known[21–23] to be associated with MDD. DIY assessment efforts also help fill a critical gap in traditional public health epidemiology research. Existing national mental health based assessments[24–27] are expensive to run and therefore infrequent by design. Some of the prior surveys are already 5-10 years old and some administered once a year and therefore miss surveying the contemporary incidence and burden of MHD.

Case B: Using real-world data to help in the early detection of suicidal ideation/attempt

Despite decades of research our ability to predict suicide has not significantly improved[28]. Less than 1% of the studies look at individualized and proximal risk factors[29]. Web-based searches are another information-rich source of passively collected data that can potentially uncover health-related behavior based on the proximity and type of information sought by the user. On average there are about 3.5 billion search queries[30] on Google every day. Of these close, to 5% (~175 million)search queries[30,31] are for seeking health-related information daily. Estimates[32] also show increasing interest in queries related to Mental Health and Depression over the last 10 years. This is highly contextual information solicited by users at the population scale and can be used to inform public health. Past research has shown the application of web-searches to identify a variety of health domains including identifying adverse drug reactions/recalls[33,34] and detecting neurodegenerative disorders[35]. With research[95] indicating that 70% of cases in the last 50 years are related to externalizing symptoms and life events; the person-generated web-search data has the potential to be used for early detection of suicidal ideation. Diagnosis of major depressive disorder is also related to a higher risk of suicide[36] with almost a 20 fold increase[37]. Research shows[38], every successful suicide takes upto 25 attempts. Despite that, our ability to predict suicide and save lives has not changed in the last 50 years[28]. Suicide rates have risen in the last 20 years[39] and so far less

than 1% of the suicide-related studies looking at proximal risk factors[29]. Investigating the utility of web-search to understand proximal risk factors has the potential to bend the curve on suicide prediction research and initiate deployment of tools that can help monitor vulnerable groups particularly ones with a prior attempt. While exploratory in nature, some of my own early work has shown the utility of individuals' web-search data to assess the proximal risk factors for suicide. Through a pilot study, I have developed a pipeline (gTAP - google takeout analysis pipeline) that is able to securely download, process and featurize real-world data from web-based searches conducted on Google search engine. The on-going pilot study has the potential to generate new evidence to help us understand novel proximal and personalized risk factors for suicide and to my knowledge has not been researched before.

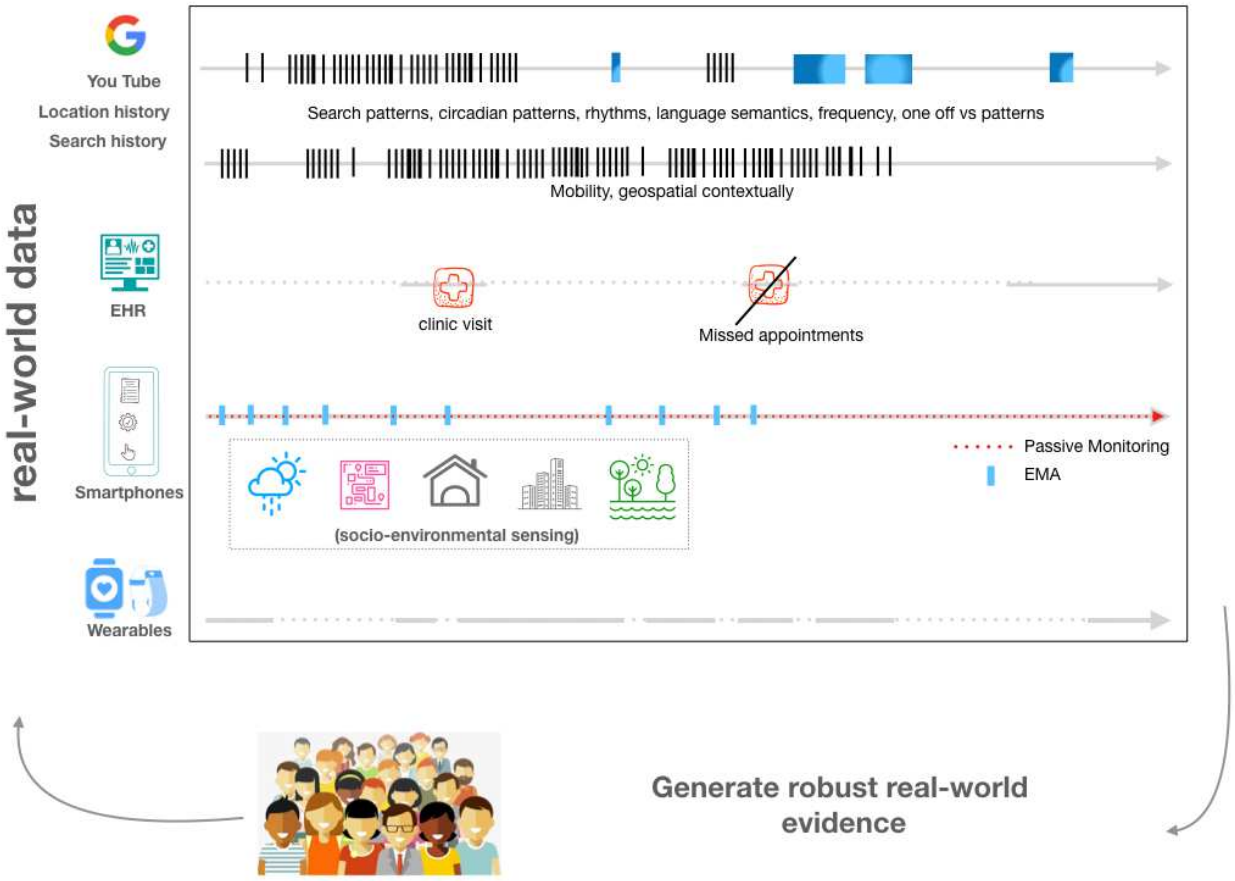


Figure 6.1. Multiple sources of real-world data (RWD) for enabling future pragmatic clinical trials.

6.2 REFERENCES

1. Suicide Statistics — AFSP. *AFSP* Available at: https://afsp.org/about-suicide/suicide-statistics/. (Accessed: 4th November 2018)

2. Yeager, C. M. & Benight, C. C. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. *Mhealth* **4**, 37 (2018).

3. Gilchrist, I. C. We Can Build It, But Will They Come? *Catheter. Cardiovasc. Interv.* **77**, 818–819 (2011).

4. Dorsey, E. R. *et al.* The Use of Smartphones for Health Research. *Acad. Med.* **92**, 157–160 (2017).

5. Kaplan, R. M., Chambers, D. A. & Glasgow, R. E. Big data and large sample size: a cautionary note on the potential for bias. *Clin. Transl. Sci.* **7**, 342–346 (2014).

6. Chaibub Neto, E. *et al.* Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *NPJ Digit Med* **2**, 99 (2019).

7. Neto, E. C. *et al.* A Permutation Approach to Assess Confounding in Machine Learning Applications for Digital Health. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19* (2019). doi:10.1145/3292500.3330903

8. Doerr, M. *et al.* Implementing a universal informed consent process for the Research Program. *Pac. Symp. Biocomput.* **24**, 427–438 (2019).

9. All of Us Research Program Investigators *et al.* The 'All of Us' Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).

10. Sanders, E. B.-N., -N. Sanders, E. B. & Stappers, P. J. Co-creation and the new landscapes of design. *Design: Critical and Primary Sources* (2016). doi:10.5040/9781474282932.0011

11. Torous, J. *et al.* Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry* **18**, 97–98 (2019).

12. Moller, A. C. *et al.* Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. *J. Behav. Med.* **40**, 85–98 (2017).

13. Keogh, A., Tully, M. A., Matthews, J. & Hurley, D. A. A review of behaviour change theories and techniques used in group based self-management programmes for chronic low back pain and arthritis. *Man. Ther.* **20**, 727–735 (2015).

14. Swift, B. *et al.* Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care. *Clin. Transl. Sci.* **11**, 450–460 (2018).

15. Ho, Y., Hu, F. & Lee, P. The Advantages and Challenges of Using Real□World Data for Patient Care. *Clinical and Translational Science* (2019). doi:10.1111/cts.12683

16. Marquis-Gravel, G. *et al.* Technology-Enabled Clinical Trials: Transforming Medical Evidence Generation. *Circulation* **140**, 1426–1436 (2019).

17. Sillence, E., Briggs, P., Harris, P. R. & Fishwick, L. How do patients evaluate and make use of online health information? *Soc. Sci. Med.* **64**, 1853–1862 (2007).

18. Majority of Adults Look Online for Health Information. *Pew Research Center* (2013). Available at: http://www.pewresearch.org/fact-tank/2013/02/01/majority-of-adults-look-online-for-health-information/. (Accessed: 4th November 2018)

19. Zhang, Y., Jones, B., Spalding, M., Young, R. & Ragain, M. Use of the Internet for Health Information Among Primary Care Patients in Rural West Texas. *South. Med. J.* **102**, 595–601 (2009).

20. Tonsaker, T., Bartlett, G. & Trpkov, C. Health information on the Internet: gold mine or minefield? *Can. Fam. Physician* **60**, 407–408 (2014).

21. Kessler, R. C. The effects of stressful life events on depression. *Annu. Rev. Psychol.* **48**, 191–214 (1997).

22. Patel, V. *et al.* Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities , 3rd edition. *Lancet* **387**, 1672–1685 (2016).

23. Li, M., D'Arcy, C. & Meng, X. Maltreatment in childhood substantially increases the risk of adult depression and anxiety in prospective cohort studies: systematic review, meta-analysis, and proportional attributable fractions. *Psychol. Med.* **46**, 717–730 (2016).

24. NHANES - National Health and Nutrition Examination Survey Homepage. (2018). Available at: https://www.cdc.gov/nchs/nhanes/index.htm. (Accessed: 4th November 2018)

25. National Comorbidity Survey. Available at: https://www.hcp.med.harvard.edu/ncs/. (Accessed: 4th November 2018)

26. Kessler, R. C. *et al.* National comorbidity survey replication adolescent supplement (NCS-A): II. Overview and design. *J. Am. Acad. Child Adolesc. Psychiatry* **48**, 380–385 (2009).

27. National Survey on Drug Use and Health. Available at: https://nsduhweb.rti.org/respweb/homepage.cfm. (Accessed: 4th November 2018)

28. Franklin, J. C. *et al.* Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychol. Bull.* **143**, 187–232 (2017).

29. Christensen, H., Cuijpers, P. & Reynolds, C. F. Changing the Direction of Suicide Prevention Research. *JAMA Psychiatry* **73**, 435 (2016).

30. Google Search Statistics - Internet Live Stats. Available at: http://www.internetlivestats.com/google-search-statistics/. (Accessed: 4th November 2018)

31. Google. A remedy for your health-related questions: health info in the Knowledge Graph. *Official Google Blog* Available at: https://googleblog.blogspot.com/2015/02/health-info-knowledge-graph.html. (Accessed: 4th November 2018)

32. Searching for Health. Available at: https://googlenewslab.gistapp.com/searching_for_health/. (Accessed: 4th November 2018)

33. Pages, A., Bondon-Guitton, E., Montastruc, J. L. & Bagheri, H. Undesirable Effects Related to Oral Antineoplastic Drugs: Comparison Between Patients' Internet Narratives and a National Pharmacovigilance Database. *Drug Saf.* **37**, 629–637 (2014).

34. Yom-Tov, E. Predicting Drug Recalls From Internet Search Engine Queries. *IEEE J Transl Eng Health Med* **5**, 4400106 (2017).

35. White, R. W., Murali Doraiswamy, P. & Horvitz, E. Detecting neurodegenerative disorders from web search signals. *npj Digital Medicine* **1**, (2018).

36. Turecki, G. & Brent, D. A. Suicide and suicidal behaviour. *Lancet* **387**, 1227–1239 (2016).

37. Chesney, E., Goodwin, G. M. & Fazel, S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. *World Psychiatry* **13**, 153–160 (2014).

38. Suicide Statistics — AFSP. *AFSP* Available at: https://afsp.org/about-suicide/suicide-statistics/. (Accessed: 4th November 2018)

39. Suicide rates rising across the U.S. | CDC Online Newsroom | CDC. (2018). Available at: https://www.cdc.gov/media/releases/2018/p0607-suicide-prevention.html. (Accessed: 5th November 2018).

APPENDIX A

Table A-1. Comparison of median survival time (with 95% confidence intervals) across studies using two analytical approaches with and without censoring.

		No censoring		With right censoring	
Strata	**N**	**median**	**95% CI**	**median**	**95% CI**
SleepHealth	12850	2	2-2	2	2-2
Brighten	875	26	17-33	26	18-33
Asthma	4666	12	11-13	12	11-13
ElevateMS	605	7	5-10	7	5-10
mPower	6908	4	4-5	4	4-4
Phendo	7802	4	3-4	3	3-4
MyHeartCounts	18671	9	9-9	8	8-9
Start	42237	2	2-2	2	2-3

Table A-2. Comparison of median survival time (with 95% confidence intervals) across gender (Female vs Male) stratified by studies using two analytical approaches with and without censoring.

		No censoring		With right censoring	
Strata	**N**	**median**	**95% CI**	**median**	**95% CI**
gender=Female|SleepHealth	3659	2	2-2	2	2-2
gender=Female|Brighten	681	25	16-33	26	16-34
gender=Female|Asthma	993	9	7-12	9	7-11
gender=Female|ElevateMS	244	35	26-49	39	26-53
gender=Female|mPower	1998	4	4-5	4	3-5
gender=Female|Phendo	7532	4	3-4	3	3-4
gender=Female|MyHeartCounts	1320	12	11-13	12	11-13
gender=Female|Start	32067	2	2-2	2	2-3
gender=Male|SleepHealth	8898	2	2-2	2	1-2
gender=Male|Brighten	194	30	12-44	30	12-44
gender=Male|Asthma	1516	7	5-9	7	6-9
gender=Male|ElevateMS	85	26	16-57	45	26-75
gender=Male|mPower	4886	4	4-5	4	4-4
gender=Male|MyHeartCounts	5641	14	13-14	14	13-15
gender=Male|Start	10170	2	2-2	3	2-3

Table A-3. Comparison of median survival time (with 95% confidence intervals) across age groups stratified by studies using two analytical approaches with and without censoring.

		No censoring		With right censoring	
Strata	**N**	**median**	**95% CI**	**median**	**95% CI**
age_group=18-29|SleepHealth	4063	2	2-2	1	1-1
age_group=18-29|Brighten	438	18.5	13-31	19.5	13-31
age_group=18-29|Asthma	1088	4	3-4	3	2-4
age_group=18-29|ElevateMS	62	1	1-3	1	1-2
age_group=18-29|mPower	2139	2	1-2	1	1-1
age_group=18-29|Phendo	4145	3	3-4	3	3-4
age_group=18-29|MyHeartCounts	390	2	2-4	2	1-3
age_group=18-29|Start	23537	2	2-2	2	2-2
age_group=30-39|SleepHealth	3558	2	2-2	2	1-2
age_group=30-39|Brighten	220	35	20-49	35	20-49
age_group=30-39|Asthma	699	8	6-10	8	6-11
age_group=30-39|ElevateMS	151	6	2-12	5	2-12
age_group=30-39|mPower	1251	3	2-3	2	2-3
age_group=30-39|Phendo	2701	4	3-5	4	3-4
age_group=30-39|MyHeartCounts	505	3	2-4	3	2-6
age_group=30-39|Start	10200	2	2-3	3	3-4
age_group=40-49|SleepHealth	2574	2	2-3	2	2-2
age_group=40-49|Brighten	129	39	12-65	39	12-68
age_group=40-49|Asthma	362	12	9-19	13	9-21
age_group=40-49|ElevateMS	162	8	5-13	7.5	5-13
age_group=40-49|mPower	898	5	4-6	5	4-6
age_group=40-49|Phendo	616	4	3-6	4	3-5
age_group=40-49|MyHeartCounts	253	8	5-10	7	4-10
age_group=40-49|Start	5236	2	2-2	3	2-3
age_group=50-59|SleepHealth	1363	3	3-4	3	2-4
age_group=50-59|Brighten	61	42	14-83	42	14-83
age_group=50-59|Asthma	228	32.5	19-53	36	20-56
age_group=50-59|ElevateMS	126	12	7-24	11	7-24
age_group=50-59|mPower	927	9	7-10	10	8-12
age_group=50-59|Phendo	19	1	1-30	1	1-73
age_group=50-59|MyHeartCounts	188	6	4-10	6	4-10
age_group=50-59|Start	2204	2	1-2	3	2-4
age_group=60+|SleepHealth	833	4	3-6	4	3-6
age_group=60+|Brighten	27	1	1-25	1	1-25
age_group=60+|Asthma	135	57	40-70	57	39-70
age_group=60+|ElevateMS	68	18	8-53	20	8-75

age_group=60+|mPower	1588	15	13-17	15	13-19
age_group=60+|Phendo	3	1	1-1	1	1-1
age_group=60+|MyHeartCounts	216	9.5	7-17	11	7-26
age_group=60+|Start	1046	1	1-1	2	1-2

Table A-4. Comparison of median survival time (with 95% confidence intervals) across disease of interest to the study(True/False) stratified by studies using two analytical approaches with and without censoring.

		No Censoring		With right censoring	
Strata	**N**	**median**	**95% CI**	**median**	**95% CI**
caseStatus=FALSE|SleepHealth	12269	2	2-2	2	1-2
caseStatus=FALSE|Asthma	86	12	6-20	12.5	5-20
caseStatus=FALSE|ElevateMS	132	1	1-2	1	1-1
caseStatus=FALSE|mPower	4049	3	3-3	3	2-3
caseStatus=FALSE|MyHeartCounts	15666	10	9-10	9	9-9
caseStatus=TRUE|SleepHealth	500	16.5	14-20	16.5	14-20
caseStatus=TRUE|Asthma	4249	14	13-15	14	13-15
caseStatus=TRUE|ElevateMS	473	12	8-14	11	7-15
caseStatus=TRUE|mPower	1896	15	13-17	15.5	14-19
caseStatus=TRUE|MyHeartCounts	1009	10	9-11	10	9-11

Table A-5. Comparison of median survival time (with 95% confidence intervals) across participants that were clinically referred (True/False) stratified by studies using two analytical approaches with and without censoring. *After right censoring, the retention in the clinically-referred cohort in the mPower study did not drop below 50%. Therefore the lowest retention time (76 days) for 52.9% retention is shown.

		No censoring		With right censoring	
Strata	**N**	**median**	**95% CI**	**median**	**95% CI**
clinicalReferral=FALSE|ElevateMS	469	4	2-6	4	2-6
clinicalReferral=FALSE|mPower	6836	4	4-4	4	4-4
clinicalReferral=TRUE|ElevateMS	136	25.5	17-55	26	17-56
clinicalReferral=TRUE|mPower	72	58.5	51-83	> 76*	55-NA

Table A-6. Change in median survival time (with 95% CI) based on the minimum number of days (N = 1-32) a subset of participants continued to use the study apps

Study	minimum duration	median	95% CI
SleepHealth	1	2	2-2
Brighten	1	26	17-33
Asthma	1	12	11-13
ElevateMS	1	7	5-10
mPower	1	4	4-5
Phendo	1	4	3-4
MyHeartCounts	1	9	9-9
Start	1	2	2-2
SleepHealth	2	6	6-7
Brighten	2	71	63-81
Asthma	2	23	21-24
ElevateMS	2	26	21-35
mPower	2	13	13-14
Phendo	2	16	15-17
MyHeartCounts	2	12	12-13
Start	2	16	16-17
SleepHealth	4	13	12-13
Brighten	4	75	66-83
Asthma	4	27	25-29
ElevateMS	4	35	26-46
mPower	4	20	19-21
Phendo	4	25	24-27
MyHeartCounts	4	14	14-14
Start	4	21	20-22
SleepHealth	8	21	20-22
Brighten	8	81	72-83
Asthma	8	33	31-35
ElevateMS	8	48	37-57

mPower	8	28	26-29
Phendo	8	36	34-38
MyHeartCounts	8	17	16-17
Start	8	29	28-29
SleepHealth	16	36	34-38
Brighten	16	83	82-83
Asthma	16	44	42-46
ElevateMS	16	69	57-75
mPower	16	42	41-44
Phendo	16	49	47-51
MyHeartCounts	16	35	34-36
Start	16	35	35-36
SleepHealth	32	60	57-62
Brighten	32	84	83-NA
Asthma	32	64	61-66
ElevateMS	32	77	75-81
mPower	32	58	56-60
Phendo	32	65	63-67
MyHeartCounts	32	60	59-62
Start	32	52	51-53

Table A-7. Proportion of missingness in the selected demographics across all the eight studies

Demographics Variable %(N)	Asthma	Brighten	ElevateMS	mPower	MyHeartCounts	Phendo	SleepHealth	Start
Age	41.24 (1763)	0.8 (7)	5.95 (36)	1.53 (106)	91.47 (16669)	0.64 (48)	2.96 (378)	0.03 (14)
Gender	41.31 (1766)	0.8 (7)	45.62 (276)	0	83.1 (15145)	0	1.66 (212)	0
Race	23.42 (1001)	0.8 (7)	44.79 (271)	0.46 (32)	74.25 (13532)	0.03 (2)	58.41 (7459)	NA
State	62.83 (2686)	7.19 (63)	45.45 (275)	7.09 (490)	41.55 (7572)	53.32 (4016)	100 (12770)	44.84 (19148)

Table A-8. Summary of participants' daily study app usage behavior and demographics across five distinct engagement clusters

Cluster Characteristics	dedicated	high	moderate	sporadic	abandoners
	C1	**C2**	**C3**	**C4**	**C5***
N	3,369	4,847	14,339	16,607	55,452
Proportion of Participants(%) (median ± IQR)	3.5 ± 2.5	6 ± 3.1	12.5 ± 2.9	21 ± 9.6	54.6 ± 17.3
Duration in study app (median ± IQR)	81 ± 22	53 ± 37	18 ± 21	19 ± 25	1 ± 1
Regularity of app usage in the first 84 days(%) (median ± IQR)	96.4 ± 26.2	63.1 ± 44.1	21.4 ± 25	22.6 ± 29.8	1.2 ± 1.2
Days Active in study app (median ± IQR)	39 ± 39	24 ± 13	9 ± 5	4 ± 3	1 ± 1
Days between activity (median ± IQR)	2 ± 1	2 ± 1	2 ± 1	5 ± 7	NA
Age group - (median ± IQR)					
18-29	20.5 ± 29.2	28.3 ± 31.7	33.7 ± 35.7	33 ± 26	45.1 ± 18.6
30-39	21.5 ± 12.4	26.5 ± 7.3	27.8 ± 4.6	28 ± 8.1	27.8 ± 7.7
40-49	15.7 ± 5.4	16 ± 9.6	14.2 ± 10.8	15.5 ± 4.3	13.6 ± 3.3
50-59	12.4 ± 12.5	12.2 ± 9.1	10 ± 8.8	11.2 ± 8.5	8 ± 4.9
60+	17.2 ± 29.3	15.1 ± 11.7	11.7 ± 11.3	6.3 ± 9.2	5.1 ± 6.6
Race/Ethnicity (median ± IQR)					
Non-Hispanic White	82.8 ± 4.8	78.4 ± 9.9	77.6 ± 9.3	71.6 ± 9.3	73.6 ± 8.5
Hispanic/Latinos	6.5 ± 3.6	7.5 ± 5.4	10.3 ± 6.6	8.8 ± 5.4	9.3 ± 8.1
African-American/Black	5.3 ± 3.5	3.3 ± 1.6	3.2 ± 2.8	5.3 ± 3.2	4.4 ± 3.5
Asian	3.7 ± 2.8	4.8 ± 2.1	5.3 ± 2.6	8 ± 3.6	6 ± 2.5
AIAN	0.7 ± 0.5	0.6 ± 0.4	0.3 ± 0.2	0.8 ± 0.1	0.4 ± 0.3
Hawaiian or other Pacific Islander	0.6 ± 0.3	0.5 ± 0.6	0.3 ± 0.7	0.6 ± 0.6	0.5 ± 0.4
Other	3.5 ± 1.8	3.8 ± 2.7	5 ± 3.1	4.5 ± 2.3	6 ± 2.6

Figure A-1. Sensitivity analysis comparing the participant retention across **a)** age groups including participants for whom age was not available and **b)** gender including participants with missing gender

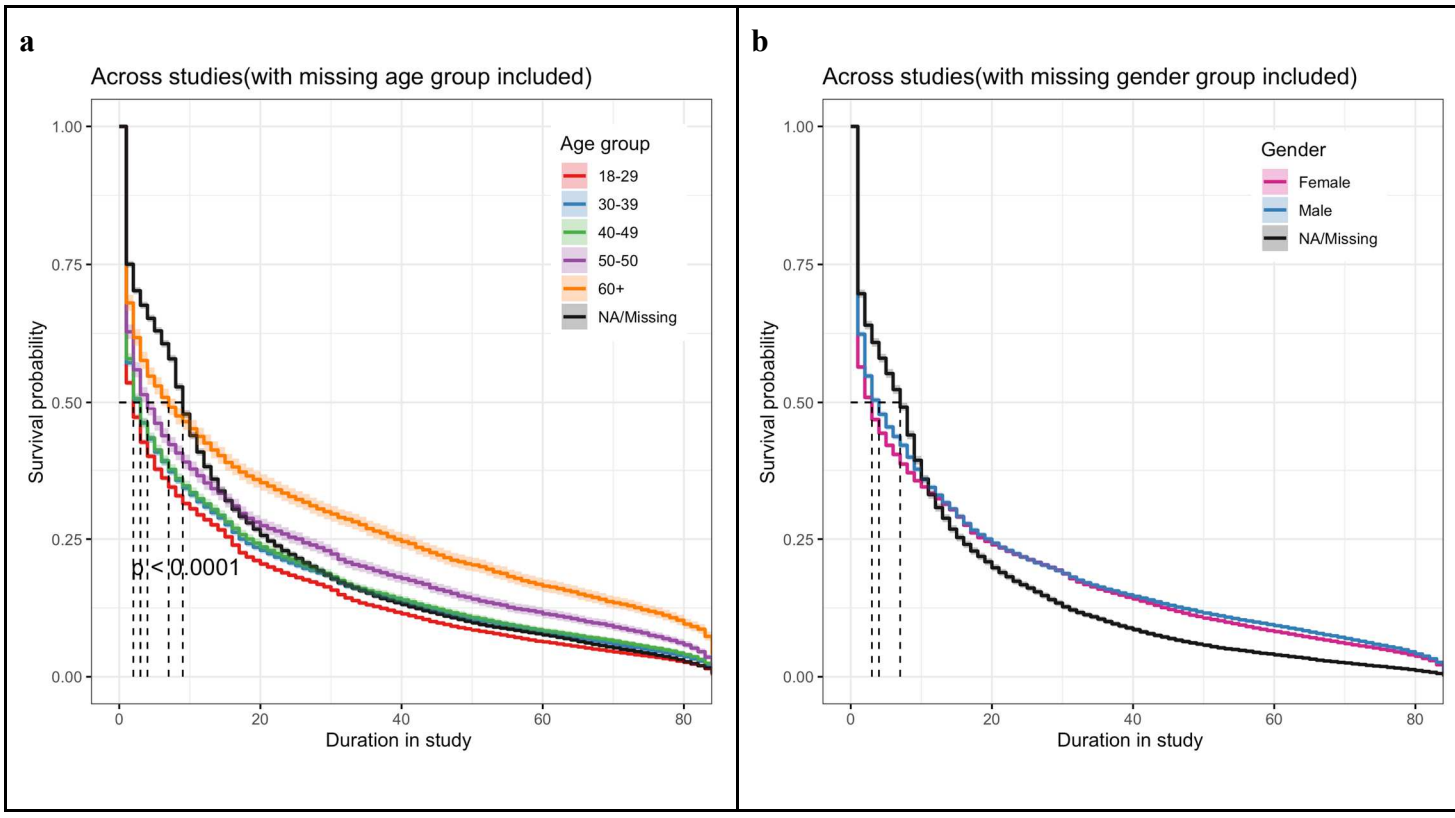
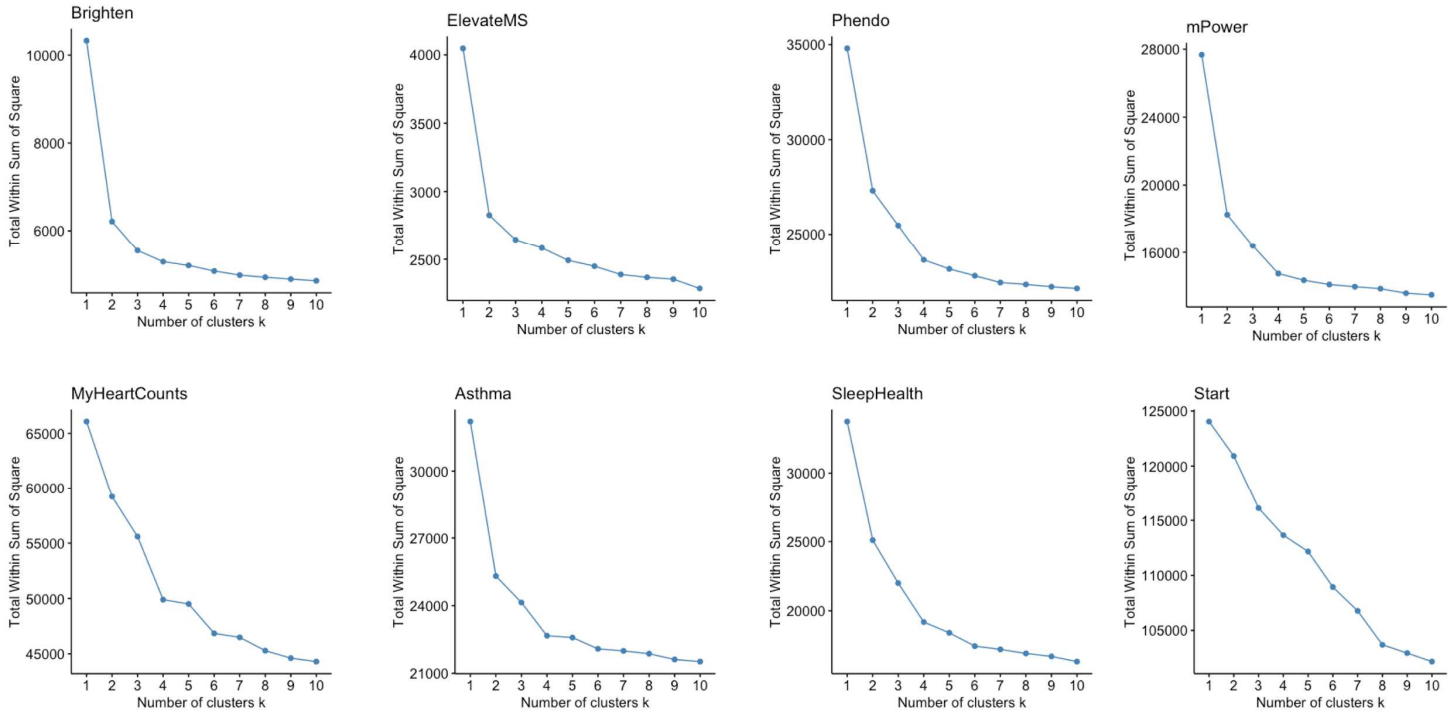


Figure A-2. Comparison of within-cluster variations across studies using different cluster size (N=1-10) for K-means clustering. Seven out of eight studies showed the within-cluster variation to be minimum by partitioning the daily app usage into four clusters.



APPENDIX B

Table B-1.Screening Questions

1. Do you use social media? ☐ Yes ☐ No
   If yes, check all that apply
   a. Facebook
   b. Snapchat
   c. Instagram
   d. Twitter
   e. Pinterest
   f. LinkedIn
   g. Tumblr
   h. Vero
   i. Other: _____(please indicate)

2. Are you 18 years of age or older? ☐ Yes ☐ No

3. Are you _____ (insert race)? ☐ Yes ☐ No
4. How old are you?

   ☐ 18-24 ☐ 25-39 ☐ 40-54 ☐ 55-69 ☐ 70+

5. I live in the United States? ☐ Yes ☐ No (if no, go to exit message)

6. I graduated from high school in the United States? ☐ Yes ☐ No

7. What is your race? (Check all that apply)
☐ Latino
☐ White
☐ Black
☐ Asian
☐ Hawaiian or Pacific Islander, Native American, Alaska Native

8. Are you Hispanic?
a. Yes
b. No

9. What is your gender?
a. Male
b. Female
c. Non-Binary

Table B-2. T1 Survey

1. Have you ever intentionally shared your social media or search data with a research team before? ☐ Yes ☐ No

2. Have you ever volunteered for a research study on-line before? ☐ Yes ☐ No

3. If you saw an ad for a pharmaceutical company sponsored research study for a health condition such as depression (for example, Pfizer for Prozac) on **Facebook**, would you be willing to participate?
   ☐ Yes ☐ No
   - Tell us more about the reason behind your answer. (Text box)

4. If you saw an ad for a pharmaceutical company sponsored research study for a health condition such as depression (for example, Pfizer for Prozac) during a **Google search**, would you be willing to participate? ☐ Yes ☐ No
   - Tell us more about the reason behind your answer. (Text box)

5. If you saw an ad for a university sponsored research study for a health condition such as depression (for example, UCLA for Prozac) on **Facebook**, would you be willing to participate?
   ☐ Yes ☐ No
   - Tell us more about the reason behind your answer. (Text box)

6. If you saw an ad for a university sponsored research study for a health condition such as depression (for example, UCLA for Prozac) during a **Google search**, would you be willing to participate? ☐ Yes ☐ No
   - Tell us more about the reason behind your answer. (Text box)

7. If you saw an ad for a federally sponsored research study for a health condition such as depression (for example, the National Institutes of Health for Prozac) on **Facebook**, would you be willing to participate? ☐ Yes ☐ No
   - Tell us more about the reason behind your answer. (Text box)

8. If you saw an ad for a federally sponsored research study for a health condition such as depression (for example, the National Institutes of Health for Prozac) during a **Google search**, would you be willing to participate? ☐ Yes ☐ No
   - Tell us more about the reason behind your answer. (Text box)

9. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with a pharmaceutical company sponsored research study for a health condition like depression (for example, Pfizer)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

10. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with a university sponsored research study for a health condition like depression (for example, the University of California)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

11. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with a federally sponsored research study for a health condition like depression (for example, the National Institutes of Health)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

12. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with a health plan sponsored research study for a health condition like depression (for example, United Health Care)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

13. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with an foundation sponsored research study for a health condition like depression (for example, the Bill & Melinda Gates Foundation)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

Table B-3. T2 Survey

1. Since the last survey, have you stopped using any particular social media?
   ☐ Yes ☐ No
   If yes, check all that apply
   a. Facebook
   b. Snapchat
   c. Instagram
   d. Twitter
   e. Pinterest
   f. LinkedIn
   g. Tumblr
   h. Vero
   i. Reddit
   j. Other: _____(please indicate)

2. Since the last survey, have you joined any particular social media platforms?
   ☐ Yes ☐ No
   If yes, check all that apply
   a. Facebook
   b. Snapchat
   c. Instagram
   d. Twitter
   e. Pinterest
   f. LinkedIn
   g. Tumblr
   h. Vero
   i. Reddit
   j. Other: _____(please indicate)

3. Have you ever intentionally shared your social media or search data with a research team before? ☐ Yes ☐ No
   If yes, check all that apply
   a. Facebook
   b. Snapchat
   c. Instagram
   d. Twitter
   e. Pinterest
   f. LinkedIn
   g. Tumblr

h. Vero
i. Reddit
j. Other: _____(please indicate)

If no, what kind of social media data would you be willing to share with a research team?

a. Facebook
b. Snapchat
c. Instagram
d. Twitter
e. Pinterest
f. LinkedIn
g. Tumblr
h. Vero
i. Reddit
j. Other: _____

4. Have you ever volunteered for a research study on-line before? ☐ Yes ☐ No

If Yes:

a. What category below matches closely to the research study you participated in? (Check all that apply) (note - allow people to click more than one)
    1. mTurk based studies
    2. Online health surveys other than mTurk
    3. Online focus groups other than mTurk
    4. Online marketing surveys other than mTurk
    5. Other online surveys
    6. University Sponsored research
    7. Pharma Sponsored research
    8. Non-profit Sponsored research
    9. Other : _____
b. Will you tell us briefly why you volunteered to participate in that online research study?
c. If no, please tell us more about the reason behind your answer.

5. Would you be willing to click an ad to learn more about a research study for a health condition based on which social media you saw the ad on? ☐ Yes ☐ No

If yes, on which social media sites would you be most willing to interact with an ad?:

a. Facebook
b. Snapchat
c. Instagram

d. Twitter
e. Pinterest
f. LinkedIn
g. Tumblr
h. Vero
i. Reddit
j. Other: _____

6. Would you be more willing to click an ad to **learn(not necessarily participate)** about an online research study for a health condition based on who is sponsoring the study? ☐ Yes ☐ No

If yes, which would you be most willing to click on?:

A study sponsored by:

a. University (e.g: University of California, Eckerds College)
b. Non-profit (e.g: Bill & Melinda Gates Foundation)
c. Federal Agency (e.g: National Institutes of Health)
d. Pharmaceutical company (e.g: Novartis, Pfizer)

7. If you saw an ad for a pharmaceutical company sponsored research study for a health condition such as depression (for example, Pfizer for Prozac) on **Facebook**, would you be willing to participate?
☐ Yes ☐ No

- Tell us more about the reason behind your answer. (Text box)

8. If you saw an ad for a pharmaceutical company sponsored research study for a health condition such as depression (for example, Pfizer for Prozac) during a **Google search**, would you be willing to participate? ☐ Yes ☐ No

- Tell us more about the reason behind your answer. (Text box)

9. If you saw an ad for a university sponsored research study for a health condition such as depression (for example, UCLA for Prozac) on **Facebook**, would you be willing to participate?
☐ Yes ☐ No

- Tell us more about the reason behind your answer. (Text box)

10. If you saw an ad for a university sponsored research study for a health condition such as depression (for example, UCLA for Prozac) during a **Google search**, would you be willing to participate? ☐ Yes ☐ No

- Tell us more about the reason behind your answer. (Text box)

11. If you saw an ad for a federally sponsored research study for a health condition such as depression (for example, the National Institutes of Health for Prozac) on **Facebook**, would you be willing to participate? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

12. If you saw an ad for a federally sponsored research study for a health condition such as depression (for example, the National Institutes of Health for Prozac) during a **Google search**, would you be willing to participate? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

13. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with a pharmaceutical company sponsored research study for a health condition like depression (for example, Pfizer)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

14. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with a university sponsored research study for a health condition like depression (for example, the University of California)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

15. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with a federally sponsored research study for a health condition like depression (for example, the National Institutes of Health)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

16. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with a health plan sponsored research study for a health condition like depression (for example, United Health Care)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

17. Would you be willing to share your social media data (Facebook, Twitter, Instagram posts) with an foundation sponsored research study for a health condition like depression (for example, the Bill & Melinda Gates Foundation)? ☐ Yes ☐ No
    - Tell us more about the reason behind your answer. (Text box)

18. Based on the results of our first survey, we noticed a higher percentage of users are willing to participate in University-sponsored research studies recruiting through Facebook ads compared to Google ads.
    a. Are you more willing to participate in a university-sponsored research study you found out about from an ad on facebook as opposed to one you saw on Google?

b. Why? Why not? (Text box)

18. Recently, social media companies have been more open about their privacy policies and their commitment to your privacy especially in the wake of new the European General Data Protection Regulation (GDPR)(link- https://www.eugdpr.org/)regulations. For example, Facebook has recently released a video (https://www.youtube.com/watch?v=Q4zd7X98eOs). Additionally, other companies sent out emails highlighting changes in their data usage and privacy policies.

19. Have you seen any data privacy and security emails and or advertisements from Facebook, Google, or other social media websites in the last 3 months? Yes / No.

Table B-4. Number and proportion of participants willing to participate seeing a study ad as part of the Google Search results

Recruitment Platform Facebook	T1 Survey						T2 Survey					
	Pharma		Federal		University		Pharma		Federal		University	
Willingness to Participate	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
N(%)	392 (42.9)	522 (57.1)	398 (43.5)	516 (56.5)	516 (56.5)	398 (43.5)	228 (34.9)	424 (65.0)	274 (42.1)	378 (57.9)	386 (59.2)	266 (40.8)
Age (%)												
18-24	38 (9.7)	38 (7.3)	41 (10.3)	35 (6.8)	48 (9.3)	28 (7.0)	21 (9.2)	30 (7.1)	22 (8.0)	29 (7.7)	28 (7.3)	23 (8.6)
25-39	237 (60.5)	291 (55.7)	234 (58.8)	294 (57.0)	315 (61.0)	213 (53.5)	119 (52.2)	259 (61.1)	161 (58.8)	217 (57.4)	220 (57.0)	158 (59.4)
40-54	96 (24.5)	130 (24.9)	93 (23.4)	133 (25.8)	121 (23.4)	105 (26.4)	69 (30.3)	96 (22.6)	64 (23.4)	101 (26.7)	102 (26.4)	63 (23.7)
55 and over	21 (5.4)	63 (12.1)	30 (7.5)	54 (10.4)	32 (6.2)	52 (13.1)	19 (8.4)	39 (9.2)	27 (9.9)	31 (8.2)	36 (9.3)	22 (8.2)
Gender												
Male (%)	174 (44.4)	246 (47.1)	191 (48.0)	229 (44.4)	244 (47.3)	176 (44.2)	97 (42.5)	211 (49.8)	130 (47.4)	178 (47.1)	177 (45.9)	131 (49.2)
Race/Ethnicity (%)												
White	260 (66.3)	355 (68.0)	271 (68.1)	344 (66.7)	353 (68.4)	262 (65.8)	149 (65.4)	288 (67.9)	184 (67.2)	253 (66.9)	266 (68.9)	171 (64.3)
Asian	21 (5.4)	31 (5.9)	22 (5.5)	30 (5.8)	28 (5.4)	24 (6.0)	13 (5.7)	27 (6.4)	16 (5.8)	24 (6.3)	20 (5.2)	20 (7.5)
Black/African American	47 (12.0)	60 (11.5)	46 (11.6)	61 (11.8)	57 (11.0)	50 (12.6)	36 (15.8)	50 (11.8)	36 (13.1)	50 (13.2)	48 (12.4)	38 (14.3)
Hawaiian/PI/NAmerican/Alaska N	9 (2.3)	4 (0.8)	6 (1.5)	7 (1.4)	8 (1.6)	5 (1.3)	3 (1.3)	5 (1.2)	1 (0.4)	7 (1.9)	2 (0.5)	6 (2.3)
Hispanic/Latino	55 (14.0)	72 (13.8)	53 (13.3)	74 (14.3)	70 (13.6)	57 (14.3)	27 (11.8)	54 (12.7)	37 (13.5)	44 (11.6)	50 (13.0)	31 (11.7)

Table B-5. Number and proportion of participants willing to participate seeing a study ad as part of the Google Search results

Recruitment Platform Google	T1 Survey						T2 Survey					
	Pharma		Federal		University		Pharma		Federal		University	
Willingness to Participate	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
N(%)	438 (47.9)	476 (52.1)	452 (49.5)	462 (50.5)	563 (61.6)	351 (38.4)	209 (32.1)	443 (67.9)	287 (44)	365 (56)	378 (57.9)	274 (42.1)
Age (%)												
18-24	48 (11.0)	28 (5.9)	48 (10.6)	28 (6.1)	59 (10.5)	17 (4.8)	18 (8.6)	33 (7.4)	23 (8.0)	28 (7.7)	28 (7.4)	23 (8.4)
25-39	248 (56.6)	280 (58.8)	263 (58.2)	265 (57.4)	327 (58.1)	201 (57.3)	113 (54.1)	265 (59.8)	164 (57.1)	214 (58.6)	212 (56.1)	166 (60.6)
40-54	111 (25.3)	115 (24.2)	102 (22.6)	124 (26.8)	132 (23.4)	94 (26.8)	59 (28.2)	106 (23.9)	72 (25.1)	93 (25.5)	100 (26.5)	65 (23.7)
55 and over	31 (7.1)	53 (11.2)	39 (7.1)	45 (9.7)	45 (8.0)	39 (11.1)	19 (9.1)	39 (8.8)	28 (9.8)	30 (8.2)	38 (10)	20 7.3)
Gender												
Male (%)	193 (44.1)	227 (47.7)	208 (46.0)	212 (45.9)	268 (47.6)	152 (43.3)	90 (43.1)	218 (49.2)	138 (48.1)	170 (46.6)	176 (46.6)	132 (48.2)
Race/Ethnicity (%)												
White/Caucasion	287 (65.5)	328 (68.9)	306 (67.7)	309 (66.9)	384 (68.2)	231 (65.8)	129 (61.7)	308 (69.5)	192 (66.9)	245 (67.1)	264 (69.8)	173 (63.1)
Asian	20 (4.6)	32 (6.7)	22 (4.9)	30 (6.5)	27 (4.8)	25 (7.1)	11 (5.3)	29 (6.5)	17 (5.9)	23 (6.3)	18 (4.8)	22 (8.0)
Black/African American	59 (13.5)	48 (10.1)	53 (11.7)	54 (11.7)	66 (11.7)	41 (11.7)	36 (17.2)	50 (11.3)	39 (13.6)	47 (12.9)	50 (13.2)	36 (13.1)
Hawaiian/PI/NAmerican/Alaska N	8 (1.8)	5 (1.1)	6 (1.3)	7 (1.5)	9 (1.6)	4 (1.1)	3 (1.4)	5 (1.1)	3 (1.0)	5 (1.4)	2 (0.5)	6 (2.2)
Hispanic/Latino	64 (14.6)	63 (13.2)	65 (14.4)	62 (13.4)	77 (13.7)	50 (14.2)	30 (14.4)	51 (11.5)	36 (12.5)	45 (12.3)	44 (11.6)	37 (13.5)

Table B-6. Reasons for agreeing or declining to participate in research studies advertised online.

	Pharma	University	Federal
Reasons to participate			
Contribute to science and help others	*"I would like to know that I could possibly help with medical research and make someone potentially feel better."*	*"I feel that a university is really trying to learn and put out good information... because they have no monetary reason to do it."*	*"People like me are a large component in helping researchers better understand conditions and diseases, so I would be willing to help."*
Trust and credibility	*"It's a pharmaceutical company doing it so they seem trustworthy in that area"*	*"I would trust university sponsored research most of all; while being targeted in the Facebook content would give me pause, I would be able to acknowledge that it is an easy way to target a specific demographic. I would expect university researchers to have fewer issues with unethical practices."*	*"I would be willing to consider participating in a government sponsored study because I would trust that it wouldn't be for profit or marketing."*
Personal or familial experience with depression	*"I would be willing because I have a family history of depression and have seen some of it's detrimental effects first-hand."*	*"I have been impacted by this ailment in the past which provides an incentive for me to help in research."*	*"I would be interested in discussing my past connection to this condition. It would be an altruistic gesture of good knowledge about my past."*
For payment	*"Pharmaceutical studies usually pay well and if the study helps me with a medical condition, that's also a benefit."*	*"I would be willing to participate based on the monetary incentive."*	*"It would likely pay very well and be professional structured."*
Reasons to not participate			

Privacy/data security concerns	*"I wouldn't want to participate in a study for a large company because I don't fully trust them to keep my information private."*	*"It would feel a little bit too private and personal."*	*"I prefer not to contribute information to the federal government database because I do not trust that the government will solely use that information for those purposes only"*
Mistrust or lack of credibility	*"I do not trust nor do I respect pharmaceutical companies. I believe there only interest is profit for themselves and not in the best interests of the public in general."*	*"I would not want to participate in a study that was not conducted by a university with a proven track record"*	*"I don't trust the government's motive behind the research."*

Table B-7. Reasons for agreeing or declining to share social media data for research studies.

	Pharma	University	Federal
Reasons to share			
Contribute to science and help others	*"because I think that even though the company is profit oriented, by sharing my posts and history maybe the company will be able to help someone upon studying my results and applying them to others."*	*"If it helped further research on something and helped people out, I probably would."*	*"Once again, I would like to help people with depression and look for alternative solutions."*
Trust and credibility	--	*"I would have more trust in the management, security, and valuing of my data with a university, which I would believe to be more motivated by knowledge than profits."*	*"They can be trusted and would likely keep all information confidential. The employees also likely have federal clearances."*
For payment	*"I would be willing, but I think the compensation would have to be really high. I'm iffy about sharing my social media data for health related reasons."*	*"I would be interested in earning money in my free time helping research."*	*"I would share my Twitter data for the right amount of money."*
Reasons not to share			
Privacy/data security concerns	*"I would not want to share my social media information with a pharmaceutical company for a research study because I wouldn't trust the company to keep my data private and safe. I would worry about having my personal information marketed to third parties."*	*"I don't care who has your data, there is NO guarantee that it will not be shared or hacked."*	*"Our privacy is invaded enough as it is by the government. I wouldn't willingly hand social media data to anything federally sponsored."*
Mistrust or lack of credibility	*"I don't trust pharmaceutical companies with my private*		*"I think they will be sneaky and not fully*

	data. They are primarily interested in profits, ethics be damned."	--	*inform me of how my data is being used."*
News of recent data breaches (e.g., Cambridge Analytica)	*"With the big thing that went on with Facebook user's information being compromised months ago, I don't feel comfortable sharing my data with anyone."*	*"I just don't share any social media data especially after the Cambridge Analytica fiasco."*	*"Probably not, unless it was completely anonymous and Facebook has had some real privacy issues lately."*

VITA

Abhishek Pratap is an academic researcher with over 10 years of full-time professional informatics experience and have worked on several digital health projects that gathered real-world evidence from thousands of participants fully remotely. He completed her PhD at the University of Washington in the Department of Biomedical Informatics and Medical Education.