UNIVERSITY OF CALIFORNIA SAN DIEGO

Leveraging Longitudinal Data for Predictive Modeling in Clinical and Mental Health Settings

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioinformatics & Systems Biology with a specialization in Biomedical Informatics

by

Argus Jerome Athanas-Crannell

Committee in charge:

Professor Nicholas J Schork, Chair
Professor Lucila Ohno-Machado, Co-Chair
Professor Hannah Carter
Professor Anders Dale
Professor Rob El-Kareh
Professor Nate Lewis

2020

The Dissertation of Argus Jerome Athanas-Crannell is approved,

and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2020

DEDICATION

Thanks to the support of my parents, Michael, Francesca, Kent, Paula, and the loving patience and keen eyes of my wife Mycalia

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Nicholas Schork as my mentor, his ability to translate complex concepts into actionable science is a skill I strive to replicate.

Chapter 1, in full, has been submitted for publication of the material as it may appear in Ophthalmology Glaucoma, 2020, Argus J Athanas, Radha Ayyagari, Linda Zangwill, Mark Christopher, Jeffrey M. Liebmann, Christopher A. Girkin, Robert M. Feldman, Harvey Dubiner, Yii-Der Chen, Kent D. Taylor, Xiuqing Guo, Jerome I. Rotter, Nicholas Schork, Robert N. Weinreb for the ADAGES III Genomics Study Group. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Journal of Medical Internet Research, 2019, Argus J Athanas, Jamison M McCorrison, Susan Smalley, Jamie Price, Jim Grady, Julie Campistron, Nicholas J Schork. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication has been submitted for publication of the material as it may appear in Journal of Medical Internet Research, 2020, Argus J Athanas, Jamison McCorrison, Julie Campiston, Nick Bender, Jamie Price, Susan Smalley, Nicholas J Schork. The dissertation author was the primary investigator and author of this paper.

VITA

2007-2011 Bachelor of Science, Chemical Engineering and Business Management, Rensselaer Polytechnic Institute

2015-2020 Doctor of Philosophy, Bioinformatics & Systems Biology with a specialization in Biomedical Informatics, University of California San Diego

PUBLICATIONS

Genome Wide Association Study and Meta-Analysis of Longitudinal Glaucoma-Related Visual Field Loss Phenotypes. *Accepted: Ophthalmology Glaucoma.* 2020

Characterizing Emotional State Transitions During Prolonged Use of a Mindfulness and Meditation App: An Observational Study. *In Review: Journal of Medical Internet Research.* 2020

Long-Term Use of a Mindfulness and Meditation Smartphone App Is Associated With Improvement in Baseline Mood, *Journal of Medical Internet Research.* 2019

ABSTRACT OF THE DISSERTATION

Leveraging Longitudinal Data for Predictive Modeling in Clinical and Mental Health Settings

by

Argus Jerome Athanas-Crannell

Doctor of Philosophy in Bioinformatics & Systems Biology

with a specialization in Biomedical Informatics

University of California San Diego, 2020

Professor Nicholas Schork, Chair

Professor Lucila Ohno-Machado, Co-Chair

We are at a pivotal point in healthcare informatics where our technical abilities and computational methods are often beyond our abilities to translate and implement them in practical and applied settings. For example, researchers are working on unraveling the intricacies of the human genome to help individuals with their healthcare needs, but often lack the ability to easily share genetic data and medical records between clinics or healthcare providers due to logistical, legal, and interoperability issues. Using more sophisticated healthcare data on individuals derived from other data-intensive assays, such imaging or digital devices, to guide their care is likely to create even more practical issues. This will especially be the case as the community pushes towards personalized medicine and an ability to track an individual's health over time through the use of these data-intensive assays. In my thesis research, I consider the application of statistical methods to large-scale longitudinal health data. I focus on not only the detection of novel relationships between longitudinal data points, but also on developing strategies for making more personalized health predictions and recommendations. I focus on three broad settings: one involving an analysis of genetic factors associated with progressive visual field loss in patients with primary open angle glaucoma using a large clinical data set, and two that consider individual mental health and mood data obtained from the prolonged use of a digital therapeutic (i.e., smartphone app) designed to provide meditations to individual users to reduce stress, anxiety, and depression. I also discuss and explore the implementation of systems designed to learn or adapt from past data in order to improve predictive models going forward using the glaucoma and healthcare app data.

INTRODUCTION

Background

Many disease risk assessments only leverage data collected on individuals at single snapshots of time.[1,2,3,4] This can be problematic for identifying factors that are truly predictive of a disease since the temporal component associated with the underlying disease process is ignored. For example, genetic studies often associate specific variants with a disease to facilitate the development of a diagnostic or enable general long-term predictions about disease risk. Without longitudinal data or very strict assumptions about disease incidence and prevalence, reliable, empirically based, long term predictions about a person's risk for developing a disease based on an individual's genetic profile are not possible. Although inherited (i.e., germline) DNA sequence -- and the genetic information within it -- doesn't change over time, it can't be used as a concrete indicator of predisposition to a disease if there is no data relating that genetic information to long term individual health outcomes. Without longitudinal and developing phenotypes data it will be difficult if not impossible know just how well genetic information may inform an individual's long-term health. Thus, while genetic association studies and methods have provided important advances in the fields of medicine and biology, they often only involve case/control or single slice in time data and thus only address part of the greater picture of how diseases manifest themselves but are limited in how they can be used to inform the prediction of individual differences in their genetically-mediated individual progression.[5]

The ease with which health data can be collected via, e.g., wireless devices, continuous glucose monitors, micro-sampling biofluids through routine dried blood spot collections, and smart phone apps, has changed the way biomedical researchers can design studies. The ubiquitous use of smart phones, mobile applications, internet-based surveys, wearables, and other remote data collection devices has enabled the gathering of aspects of this data, particularly in longitudinal contexts. In addition to the availability of continuous data collection, there has been great interest in leveraging medical records to enable the aggregated, real-time analyses for identifying trends in those records that could guide healthcare. For

example, the 2009 'HITECH' initiative was organized to motivate the 'digitalization' of all medical records and ultimately help create an environment where multi-factor, individualized healthcare could become a reality. With massive amounts of longitudinal data, researchers can look for patterns in those data that may be indicative of disease onset, mitigation of symptoms, or correlations with various factors, such as diet, that might shed light on disease pathogenesis. In addition, longitudinal data can be used to *refine* prediction models that may have clinical and public health utility. For example, with longitudinal data on biomarker or potential predictor of a particular disease one could ask how much earlier than the disease onset could one have anticipated the disease: 10 years prior, 5 years prior, 1 year prior? In addition, with longitudinal data one could ask if the use of the lifetime risk of disease determined by genetic profile (a 'trait' measure), coupled with a biomarker collected over time (a 'state' measure), could result in more reliable and precise risk mitigation strategies. Finally, longitudinal data could reveal specific trends associated with the use of a disease risk mitigation strategies that identify the settings in which those strategies might be most effective.

As a result of this ability to capture information relevant to an individual's health, there is considerable interest in personalizing medicine. Personalized medicine involves providing medical interventions – including those designed to prevent disease – based on an individual's genetic, biochemical, physiologic, exposure, and behavioral profile. One vision of personalized healthcare involves consideration of all a target patient's health data and measures and compares those data to data obtained on similar patients who may have been prescribed drugs or developed an outcome that could inform the prognosis of the target patient. The use of data in this way could then feedback into the development of predictive models that could then be used to guide patient care. These models can then be updated as experience with them is built up. This 'Rapid Learning System (RLS)' 'analyze-predict-evaluate-refine' concept can not only to be used to predict outcomes based on interventions, but also elucidate the key triggers that elicit or confound an intervention response.

There are many challenges to the analysis and use of longitudinal data, however, and I comment on these challenges in isolation in the sections that follow. First, there are a wide variety examples in the biomedical research literature that expose issues in the analysis of longitudinal data. Second, there are a number of analysis methods for longitudinal data, each of which make different assumptions, but all must deal with phenomena common to all longitudinal data, such as serial correlation, vulnerability to a wide variety of factors (such as diet, weather, etc.); 3. Third, analysis of longitudinal can lead to more powerful analyses than single-slice in time data, but only under certain conditions; 4. The general availability and use of health data collection devices, like fitbits or smart phone apps, has created vast data repositories that can be mined for patterns of relevance to clinical care and public health. However, such 'real world' data often present data analysis challenges given that there might be biases in who can take advantage of relevant data collection devices, missing data, non-uniform data collections, and other challenges; and 5. The continued analysis of accumulating longitudinal data could be exploited in efforts to improve prediction models in real time, possibly leading to the construction of 'learning systems,' which are designed to specifically evolve a system towards optimality, but themselves can be tricky to implement. Ultimately, the material in these sections provide context for the studies I pursued.

Example Longitudinal Data Analysis

The use of longitudinal data to make better clinical predictions is not a new concept, but practical applications are only now beginning to emerge. There are a few specific settings which have shown promise in the use of longitudinal data to enable more sophisticated analyses and prediction models. We provide a few examples that bring in elements of data analysis similar to those I dealt with in my research. For example, models for cardiovascular events prediction have been pursued, where researchers used EMR data and genetic information to enhance the predictive ability of those events [6]. The researchers showed further that their approach with longitudinal data performed better than those currently used in clinical practice. The researchers incorporated several different types of machine learning methods, and built predictive

models including longitudinal data and longitudinal data plus genetics. Even without the inclusion of genetic information, the models which incorporated longitudinal components outperformed the clinical standards. Genetic features marginally improved the predictive ability. Another study used EMR data and natural language processing from clinician notes to predict suicidal behavior and tendencies.[7] In this study they used a naive Bayesian classifier to model the probability of a suicidal event. Given the large sample size (> 1.7 Million individuals) and longitudinal data they were able to include a large number of covariates (>30) as well as ICD9 codes to look for predictors of suicidal events. While there was no comparison group, the researchers demonstrated an ability to predict nearly half of all suicides and suicidal behaviors an average of 3-4 years in advance with 90% specificity. Another group used temporal EMR data to model progression of kidney function loss. In their analysis the researchers derived 3 models, one which did not include longitudinal data, and the other two which did.[8] Their results showed that both temporal models vastly outperformed the model which did not include longitudinal data, at predicting kidney function loss. In this example they used generalized linear mixed-effects models to capture the non-independence between samples and account for non-linear features.

Longitudinal Data Analysis Methods

The analysis of longitudinal datasets requires special techniques, and there are several statistical frameworks that can be implemented for particular analyses.[9,10] A classic analysis method is repeated measures (rm) analysis of variance (ANOVA). This method accounts for the repeated measures on individual subjects over time, and because the analysis is fundamentally similar to standard ANOVA the results are easily understood. However, this method is not always well suited, especially for RWD, since rmANOVA operates under the assumption that variances across groups are relatively equal, and if this assumption is not met, resulting confidence intervals and p-values are biased and will lead to false positive and negative inferences. Additionally, rmANOVA assumes that the time intervals are uniform between measurements, which is rarely the case for real world data (RWD), and it does not work with unmatched

time-series data or and in situations where there is lot of missing data. There are simple alternatives to rmANOVA, but they also make assumptions. For example, if the hypothesis of interest is not concerned with the magnitude of the differences of a measurement between two or more groups, the Friedman test can be used as a non-parametric substitute to rmANOVA, but it still lacks the ability to handle asynchronous measures. Linear and non-linear mixed effects models have been a staple in handling longitudinal data, as they do not depend on many underlying assumptions and are generally more flexible than other approaches. They can also handle serial correlation (correlations between measurements collected over time on the same unit of observation) which is very likely to occur with longitudinal data. When applying mixed effects models, the only basic assumptions are that residual values follow an identifiable and specifiable distribution, like multivariate normality, and this assumption can be tested by exploring the distribution of residual values. Non-linear relationships among independent and dependent variables, as well as non-linear relationships between time and the independent and dependent variables can be handled by either transforming data prior to analysis or using a model that specifies the relationship between variables and time. This includes analysis of binary outcomes (e.g., diagnosis of disease yes or no) which can be accommodated by specifying an appropriate link function shaping the models. There are alternative statistical models that can account for longitudinal data, many of which are only now being used more and more often, but many are computationally burdensome, such as longitudinal support vector machines, and long short-term (LST) memory convolutional neural networks. These methods work well with high dimensional data but can be difficult to interpret. Given their flexibility, mixed models can be a first choice in many analyses of longitudinal data sets.

Longitudinal Data and Statistical Power

There is little doubt that leveraging longitudinal data can lead to increases in the statistical power to detect effects, but one needs to accommodate potential noise in such data, and, in hypothesis testing contexts, the assumed magnitude of an effect and number of units of observation in addition to the number of time points at which data are collection on those units of observation in the dataset. It is well known that

the inclusion of multiple measurements on each sample or unit of observation can lead to variance shrinkage of any estimated parameters, leading to greater power to test hypotheses based on those parameters.[11-13] In addition, estimates for individual subject or unit of observation variances can be made and used to account for subject level differences in those variances, which could be crucial if a parameter of interest is known to be affected by subject-specific covariates. In many translational settings the endpoints of interest are often linked to or directly correlated with an overall rate of change (such as the rate of visual loss in glaucoma, the rate of recovery from a surgery, the rate of aging, etc.). Such endpoints can only be assessed with longitudinal data.

Real World Data

Real World Data (RWD) is just as it sounds: data collected from routine use of a product and not in the context of a highly contrived or controlled setting, like a clinical trial. RWD can also be gleaned and analyzed from consumer health products that either do not need FDA approval (like dietary supplements and many smart phone health apps) or have been approved previously and are out on the market. A major component of my research involved not only using RWD longitudinal data but also considering the application of the results to guide real world use (i.e., in overt translational settings). The use RWD has presents a number of opportunities, and is actually contributing to new sources of evidence for clinical use and consumer utility.[14,15] One of the most common uses of RWD is in post-approval safety of a product, as highlighted by the FDA Sentinel Initiative to monitor long-term effects of pharmaceutical drug use.[16] This practice has been adopted by pharmaceutical and biotech companies as well to monitor safety and efficacy after Phase III drug approval and distribution in the health care market. Another successful use of RWD has been with initial regulatory approval. This has been especially true in the context of studies of rare or deadly diseases where it is difficult to obtain controls needed for randomized clinical trials (RCTs) or where it would be unethical to treat with placebo.[17,18] Thus, single arm experimental trials considering only the responses of a few patients compared to historical RWD-based controls, potentially mined from EMRs to mimic the population which is receiving the drug/treatment. In a similar vein, RWD can be used for

comparative effectiveness evaluations when there are many competing therapies and a full RTC across all compounds would be impractical. Another advantage of using RWD is that it can be used to continuously test for efficacy of treatments in the population at large. This can provide opportunities to look for subpopulations in which a particular treatment is more effective.[19] This also means that effectiveness can be evaluated on broader populations that are not confined to the strict enrollment criteria associated with an RTC.[20,21] Additionally, RWD can be used to test post-hoc alternative end points for their association with an intervention. A good example of this involves a study in which immunotherapy-treated cancer patients saw little tumor reduction during treatment and therefore were moved to different treatments. A post-hoc meta-analysis months to years after initial immunotherapy treatment showed that those pre-treated with immunotherapies had overall better survival than those that did not.[22] From a practical standpoint RWD also has the advantage that it can inform clinical decision support, quality of practice, and continually establish efficacy. This means there are likely massive cost and time savings to be reaped from leveraging RWD as long as relevant data can be cleaned and harmonized, which could help alleviate the US's inflated cost of healthcare. RWD is not without challenges - unlike the controlled setting of RTC data collection is irregular, non-standardized, contains missingness and confounding factors. Fortunately, there are solutions which can often be remedied with collection and careful incorporation of additional meta-data in relevant analyses.

Rapid Learning Systems

The development of Rapid Learning Systems (RLS), as noted, relies on a few key concepts, the most important of which involves building a longitudinal record of, e.g., patients and their outcomes, or consumers tracked for their responses or utilization of a particular product. Unfortunately, leveraging health care data is a difficult task in the US as there is no universal patient identifier and strict laws exist on handling and releasing of Patient Health Info (PHI). This issue is even more pronounced when considerations involve the use of consumer product use data, where there are even fewer standards that can be used harmonize or combine data. In terms of clinical data, it is not uncommon for a patient to receive

care from disparate sources, and without a common unique identifier, consolidating their electronic medical records (EMRs) relies on PHI and other personal records (social security number, name, address, etc.). Without this step, longitudinal records would contain duplicate entries for individuals and confound any statistical results from the data source. The CancerLinQ initiative is an example of an attempt to aggregate clinical data, and publications associated with it describe merging EMR data in detail, and show how they can use PHI to validate identities and assign quality metrics to data collected.[23,24] Another concept or component of developing an effective RLS is *tangential* data collection, in which new data types are brought into a system in order to go beyond limitations of an initial RLS. As an example, in systems designed for, e.g., precision oncology in which drugs are matched to tumor profiles, initial designs attempted to provide treatment based only on the primary tissue of the tumor. After several revisions of this approach, evidence supported a broader view in which genomics, proteomics, and metabolomics data about an individual's tumor provided better predictions for patient-drug response.[25] Of course, the use of the newer or tangential data also allows for further associations between disease and the mechanism of action and various perturbations in the underlying data which can aid in drug design, or provide evidence for additional drug indications.[26] A final concept or consideration in the development of RLS involves the addition of outcome measures that are collected in real time whereby individuals can be re-assessed, e.g., for new treatment responses if there is no evidence that a specific treatment they have been on in the past seems to have worked for them. This creates a need to continually refine the model with data that is not collected at uniform time intervals or even with the same instrument, likely adding noise to the data.

Thesis Aims

In this light, it is clear that research designed to personalize disease predictions will need to leverage data and methods that accommodate and build off longitudinal data. Longitudinal data can be exploited in a wide variety of healthcare settings, including: 1. developing insights into disease prognosis and progression; 2. increasing the power to detect important relationships between variables that capture disease relevant processes that manifest within individuals but not necessarily between individuals; and 3.

identifying and accounting for individual differences in basal or initial states that impact longitudinal outcomes (as in aggregated n-of-1 studies). Machine learning (ML) techniques are now being implemented in many settings (academic and in industry) and have amazing power to create accurate predictions as long as they have enough training data to work from. They are even being used more often to make predictions in healthcare settings[27]; however, ML techniques often suffer from becoming 'a black box.' It can be difficult to understand why they make the predictions they do, or what is driving a particular prediction, which limits their usage and adoption in real world clinical settings.[28]

Note that there is not always motivation to understand the driving features behind a predictive model if the predictions it generates are reliable. In many healthcare settings, such as developing drugs, developing infrastructure to allowing preemptive disease risk mitigation measures, and creating long-term plans for an individual's care, an understanding of the mechanisms responsible or behind the predictions is crucial since they may bear on genetic network, biochemical pathways and pathophysiological processes that could help refine future interventions and risk mitigation plans. Thus, understanding the underlying biology and causal mechanisms behind healthcare predictions can be crucial, unlike many other situations (e.g., credit card use patterns or buying habits of consumers to develop more efficient marketing strategies). Furthermore, it is unlikely that patient care will be completely automated or that this would be the ultimate goal, and to get buy-in from healthcare professionals they will probably need to understand why a certain prediction or recommendation is being made. Given these sensitivities, the majority of the analyses and model building I have pursued in this thesis to create predictions in a number of settings will complement contemporary ML techniques with various standard and advanced regression techniques. Many of these techniques have the benefit of being easy to understand with intuitive ways of interpreting the manner which predictive factors impact outputs. For example, Linear Mixed Effect (LME) models and Generalized Linear Models (GLM) are well-suited for longitudinal analysis as they can create hierarchical dependencies to account for non-independent repeated measures within a group (i.e. subject). They also can account for

interaction terms, random effects, and serial correlation, but the manner in which they are designed and their results interpreted is much more intuitive than ML-based techniques.

While there are obvious advantages to leveraging longitudinal RWD data, there are many complications that occur when obtaining, processing, and analyzing such data. For example, acquiring quality datasets that can be used to explore the benefits and issues associated with longitudinal data to develop predictive models can be complicated. Fortunately, I have been able to collaborate with, and gain access to, data from several fruitful partnerships. These include a collaboration with the Shiley Eye Center on studies of glaucoma risk and progression using longitudinal clinical information on individuals with genotype information obtained from whole genome sequencing (WGS) protocols; and a collaboration with Stop Breathe Think (SBT), Inc., a company that has collected data on 11,000,000 uses of a stress-relief mindfulness app which has been used routinely and repeatedly by over 80,000 individuals. Both the Shiley and SBT datasets are unlike many curated research datasets in that they are ongoing, observational real-world datasets. Real-world settings are different from clinical trials and laboratory tests in many ways which make processing and analyzing more challenging. For example, for the Shiley data sets, there were no placebo or contrived control groups to be used for comparison purposes in our analyses, since all patients were undergoing routine clinic care. Thus, the care they received was delivered without bias and without use of non-conventional methods, unlike in a clinical trial where the use of controls and highly contrived or specialized protocols are pursued to maximize the power to evaluate a very particular hypothesis. Real world data sets, like the Shiley dataset, include patients that have not been held on the same treatment, especially if there were no signs of the efficacy of that treatment, or there were signs of severe side effects. In addition, there were no regular scheduled intervals for testing or treatment changes since it was often up to the patients as to when they chose to receive care. Lastly, recorded outcome measures in the Shiley data set was limited by patient engagement and adherence. All of these issues can confound or conceal true effects of intervention and make finding meaningful results difficult. This was also true for the SBT data

set, as it was collected during the routine use of the SBT app and not as part of a specific study focusing on one or another element of the SBT platform.

The first chapter will address the work I pursued with the glaucoma ADAGES III[29] dataset from Shiley Eye Center. The primary outcome of glaucoma is partial vision loss or complete blindness. Glaucoma is a complex disorder and may have several underlying determinants but is primarily characterized as a neurogenerative disease which causes loss of retinal ganglion cells. Changes to the retinal nerve fiber layer, as well as the optic nerve, are also common, and can usually only be diagnosed by inspection of eye and the optic nerve. The degradation of the visual field in glaucoma is often gradual and begins at a patient's periphery, making it hard for even the patient to notice the loss immediately, even though the effects are irreversible. For this reason, early detection of the disease is important as treatment can help preserve a patient's vision. Early detection by itself is not enough though, as patients experience a wide range in rates of visual field loss and degradation, which suggests that there may be additional factors, including genetic, that mediate the disease progression. Fully understanding these factors can help improve a patient's quality of life through risk assessment, risk mitigation, early detection and treatment and potentially illuminate causal determinants which reveal novel drug targets. Ultimately, in order to understand how genetic factors impact visual loss progression, and not just glaucoma diagnosis, longitudinal data on patients with genetic data is required.

Glaucoma has been the focus of many genome wide association studies (GWAS). These studies have revealed associations between many common genetic variants and, for example, at least one form of Gluacoma, Primary Open Angel Glaucoma (POAG).[30,31,32,33] Many studies have also linked the loci harboring these variants and others to secondary phenotypes associated with the disease, such as intraocular pressure, retinal nerve fiber layer thickness, and cup to disk ratio.[34,35,36] However, there are few studies which have examined the relationship between genetic variants and glaucoma prognosis, probably because of a lack of longitudinal data. At the time of my research only one other GWAS study managed to address

the question of whether or not genetic variants impact glaucoma progress, and the authors of this study binned patients into binary classification of progressors vs. non-progressors, but did not consider progression as providing a range of values as a quantitative measure.[37] Using GWAS to find genetically-mediated progression-related associations is not uncommon, however, as GWAS on longitudinal data has been pursued with several other diseases.[38,39,40] In the analysis I have pursued, I treat visual field loss associated with glaucoma as a quantitative measure, which may reveal novel associations of relevance for predicting and treating aspects of glaucoma.

The second and third chapters will consider two different analyses involving longitudinal data arising out of the SBT data collections. These data collections include SBT app users' emotions, general mental and physical health, demographics, and the mindfulness and mediation activities they completed. These analyses explore long term changes and the effects of specific mediations provided to a user of the app for improving or maintaining their mental health. It is well-known that poor mental health, like behavioral conditions and neuropsychiatric diseases, is a stressor linked to many co-morbidities as well as low productivity.[41,42,43] Several studies suggest these conditions are on the rise, especially in the United States, yet mental health is often viewed as a stigma instead of part of something to consider in evaluating an individual's overall health.[44]

Unfortunately, even if a patient were ready to seek help for their mental health problems, and were willing to be compliant in treatment regimes, the scarcity of mental health professionals, socio-economic factors, and lack of insurance coverage for many mental health issues, restrict those who can receive care. Delivering care at the scale that is required given the seriousness and prevalence of mental health conditions, is itself a serious problem. Fortunately, there are initiatives at the FDA and other regulatory bodies to consider digital therapeutics – therapeutics delivered through a digital device such as a smartphone – bona fide, insurance-reimbursable, officially sanctioned and registered, healthcare devices. Digital therapies have the advantage of being able to scale to meet the needs of the population without increasing

the number of actual healthcare professionals. They also have the ability to reach remote locations, and act as a less expensive treatment option for those with little financial means. While digital therapeutic treatment may not be an effective tool for everyone, even if they are able to treat a fraction of people who need mental healthcare, they will have a huge impact.

Stop Breathe Think, Inc. (SBT) has created a platform that uses guided meditations to provide self-awareness and mindfulness coaching to interested users. Mindfulness training has been shown to be an effective way to treat many conditions, often inducing positive thought patterns and improving mood.[45,46] The SBT application is unique in that it asks users to define their emotional state before and after completing an meditation-based mindfulness activity. The data from the SBT device allowed me to analyze the actual effect that completing a meditation/mindfulness activity had on a user.

In chapter 2, I discuss measuring SBT app users' baseline (i.e., pre-meditation) emotional state to see if the prolonged use of the app has effects on the baseline emotional state in the long term. In chapter 3, I explore the effect of specific meditation/mindfulness activities given a user's initial emotional state given information on their background (e.g., demography). Additionally, I explore the possibility of predicting what emotional state a user might end up in if they choose certain meditations given their background, with the ultimate goal of trying to provide more personalized meditation/mindfulness activity recommendations for positively adjusting their emotional state.

Ultimately, my thesis research focuses on the application of longitudinal analysis methods using real-world data sets in order to make predictions about health outcomes that are tailored or personalized to an individual patient or health app user. I use data from one clinical care unit in a hospital and one digital device platform to showcase how one can find associations and build personalized predictive models in different settings. In addition, I show in each setting that resulting predictive models can be progressively enhanced by adding a learning component to their construction that accommodates the collection of new

data that is fed back into the predictive models. I also show that analyzing real-world longitudinal data to develop such models requires dealing with irregular measurement intervals, addressing serial correlation between observations, handling missing data, integrating disparate data types, and accommodating several additional confounding factors.

CHAPTER 1 : GENOME WIDE ASSOCIATION STUDY AND META-ANALYSIS OF LONGITUDINAL GLAUCOMA-RELATED VISUAL FIELD LOSS PHENOTYPES

Abstract

Objective: Although a great deal is known about genetic susceptibility to, and general risk factors for, glaucoma, few studies have considered the influence of genetic factors on the time course or trajectory of vision loss and related phenotypes associated with glaucoma. We therefore pursued a series of GWAS studies and meta-analyses to uncover genetic variants associated with disease progression as measured by visual loss over time. The results of our study may facilitate the selection of preventive interventions and therapies for clinical subtypes of glaucoma.

Design: Genome-wide association study (GWAS) of longitudinal visual field data followed by gene based meta-analysis of two populations with different ancestral backgrounds.

Subjects: Glaucoma patients (n = 754) with follow-up clinical data collected over several years.

Methods: We leveraged whole-genome sequence (WGS) data on 754 patients who had also been tracked for vision loss and progression over an average of 9.1 years. Focusing on the common variants from the WGS data, we pursued genome-wide association studies (GWAS) on vision loss trajectories in two sets of individuals, one of European and one of African American ancestry, using standard single locus analyses, gene-based variant analyses based on summary statistics, and a meta-analysis of the results of each study.

Outcome Measures: Glaucoma-related visual field loss trajectories collected over time

Results: We found evidence for association between variants in the TGFBR3 (p=9.44E-05), CDKN2B-AS1 (EUR: p =6.33e-06), TLR4 (EUR: p=6.43E-07, AAM: p=3.67E-05), ARHGEF12 (EUR: p=4.42E-05), TMTC2 (EUR: p=1.87E-04, AAM: p=6.86E-06), and TJP1 (AAM: 7.26E-06) genes with longitudinal visual loss phenotypes. These genes have been previously shown to harbor variants that are associated with glaucoma onset, intraocular pressure (IOP), and central corneal thickness (CCT). We also found evidence for association between variants in the region between the CNTNAP2 gene and the C7orf33 genes (p=8.984e-07, p=1.67E-06, respectively), with multiple glaucoma phenotypes.

Conclusion: The analysis of a unique and deeply-phenotyped cohort confirmed previous genetic associations with glaucoma and revealed novel associations between genetic variants and glaucoma-related progressive visual loss. Further validation of our findings is needed before our results can be used to identify glaucoma subtypes that might benefit from different intervention strategies.

Background

Glaucoma is a complex neurodegenerative disease that affects vision via detrimental changes to the optic nerve, the retinal nerve fiber layer, and retinal ganglion cell (RGC) loss. The degradation of RGC results in a narrowing of the visual field (VF) and poorer vision-related quality of life. This degradation is irreversible and can ultimately result in total blindness. Early diagnosis can help preserve a patient's vision. However, individuals with glaucoma exhibit wide variation in their rates of VF loss. Understanding the underlying mechanisms that contribute to this variation in VF loss, especially genetically-mediated mechanisms, could lead to more appropriate 'personalized' care plans for individuals suffering from, or susceptible to, glaucoma.

Previous genome wide association studies (GWAS) focused on glaucoma have established that one form of glaucoma, Primary Open Angle Glaucoma (POAG), is influenced by number of genetic variants.[5,31,32,33,47,48] Additional studies have linked genetic variants to other features of glaucoma pathology, such

as intraocular pressure, cup disk ratio, and RNFL thickness. [34,35,36] Although these studies have revealed a great deal about glaucoma's underlying genetically-mediated pathology, they do not necessarily consider factors that might be more directly relevant to the management of a glaucoma patient. For example, little research has been conducted addressing genetic determinants of glaucoma progression, or the trajectory of VF loss. Knowing the factors that may contribute to variation in VF loss over time could lead to a better understanding why standard of care is ineffective for particular patients. A study by Trikha et al, which was rooted in a GWAS with VF Progression as an end point, did consider longitudinal VF loss, but their analyses focused on a binary classification of individual progression and did not accommodate the length of follow-up, nor all the data collected over time on each patient.[37] The lack of studies on longitudinal phenotypes of relevance to glaucoma is unfortunate since GWAS have been pursued to in an effort to identify loci associated with the longitudinal trajectories and progression of many other diseases. [39,40,38] In order to overcome this, we pursued a meta-GWAS meant to identify genetic variants contributing to the progression for glaucoma-related VF loss using extensive longitudinal data on 754 patients collected, on average, over 9.1 years.

Methods

Regulatory Oversight

UCSD Internal Review Board (IRB) approval, as well as written and informed consent from patients, was obtained for this study (IRB Protocol #140107). All protocols adhered to the guidelines of the Declaration of Helsinki.

Study Population

The population we used in our analyses is a subset of the ADAGES III (clinical trial #: NCT00221923) for which we pursued WGS.[29] Patients with POAG and control participants were recruited who self-reported as African ancestry or European ancestry (note: we ultimately determined ancestry using genetic variants, see below). Eligibility for inclusion as a POAG patient required that no other ocular or

non-ocular disease was responsible for the glaucomatous VF damage. We conducted WGS on 1,589 individual patients at an average read depth of 30x using the Illumina X10 technology. European and African ancestry proportions were computed from genotype profiles using a panel of 100,000 ancestry informative markers (AIMs). After filtering for sequence quality (e.g., TS/TV ratio, SNP & INDEL counts, etc.), 1,219 Primary Open Angle Glaucoma (POAG) patients remained in the analysis. These 1,219 patients included 580 individuals who were of European descent and 639 individuals who were of African descent. We focused our analyses on a subset of glaucoma patients (n=754) who had sufficient longitudinal VF measures available. A summary of the study population is provided in Table 1.1 and the workflow is shown in Figure 1.1.

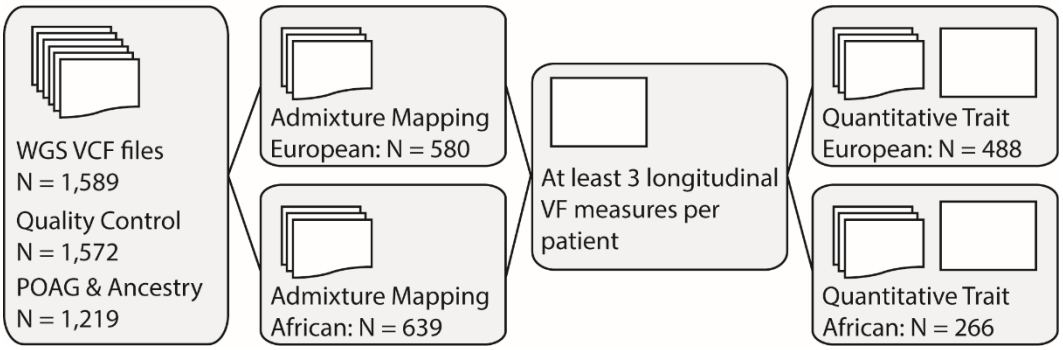


Figure 1.1 Workflow of Data Preparation
The above workflow shows various filters and divisions within the WGS dataset. Of the original 1,589 samples 17 are removed for sequence quality or identity issues, 1,219 remain after admixture filtering, and a total of 754 patients had sufficient longitudinal data.

Table 1.1 Study Population Breakdown by Ancestry
A total of 754 participants were included based on all layers of filtering. Patients were treated and studied over an average of 9.1 years. There were more Europeans, and more females included in the study.

Variable	Total	European Ancestry	African Ancestry
Count	754	488	266
Male	337	229	108
Female	417	259	158
Num VF	13.1 ± 7.5	13.6 ± 7.5	12.3 ± 7.5
Years of Follow-up	9.1 ± 4.3	9.3 ± 4.5	9.6 ± 4.0
Min	0.3	0.4	0.3
Max	18.8	18.8	18.4

Variant Calling

Variant calling was conducted by Human Longevity, Inc, using an extended version of Illumina's 'Issac' pipeline.[49] Variants were filtered using both Bcftools and PLINK v1.90.[50] Principal components analysis (PCA) was conducted to identify evidence for population stratification, and the identified PCs were later used to account for ancestry and batch effects in the association analyses. The first three PCs accounted for minimal variation but were likely reflecting minor sequencing batch effects. Roughly six million SNPs were selected for analysis based on filters for study population minor allele frequency (MAF > 5%), Hardy-Weinberg equilibrium (HWE < 1E-10), and total called genotypes across all loci interrogated (>90%) in both populations combined. Additional individuals were removed from analyses for having abnormal transversion-transition ratios, SNP/indel counts, or low confidence that their samples were from an identified patient (mis-matched sex, ancestry discordance, duplicates).

Phenotype Data

Patients underwent VF testing using the SITA-standard 24-2 strategy within 30 days from image acquisition and diagnostic tests. The UCSD Visual Field Assessment Center (VisFACT) completed a quality review of the VFs. VFs were excluded if fixation losses, false positive errors or false-negative errors

were greater than 33%. The exception was the inclusion of VFs with false-negative errors of more than 33% in eyes with advanced disease (Mean Deviation (MD) lower than 12 dB). VFs were further reviewed and excluded for the following artifacts: learning effects (i.e., initial tests showing consistent improvement on VF indexes); inappropriate fixation; inattention; lid and rim artifacts; fatigue effects; or evidence that the VF results were due to a disease other than glaucoma (such as homonymous hemianopia).

We ultimately focused on fourteen quantitative VF measures collected on the patients with the WGS data. These measures represent a patient's ability to detect light at specific thresholds and include: 1. VF global measures, Mean Deviation (MD) and Pattern Standard Deviation (PSD); 2. VF sectoral measures of mean sensitivity and pattern deviation each calculated for the 6 Garway-Heath visual field (GHVF) sectors; and 3. number of test locations for each GHVF outside normal limits (<0.05) for pattern deviation. Each of these fourteen longitudinal end points were each collapsed into progression metrics for each patient using least squares regression with visual measures as dependent variables and time since first visit as an independent variable along with covariate information (age, sex, average IOP). The slopes obtained from these regression analyses were taken as a rate of glaucoma progression. [51,52] We will use each of these phenotypes in their own GWAS analysis for a few reasons: 1) loss of VF is not uniform across the eye, and we want to be-able to detect any change as it is important for patient quality of life and 2) and SNPs can effect structural changes which may influence regional decay.[53] Because the regression model used to conduct association analyses (see below) is particularly susceptible to influence from extreme values, small sample sizes and the influence of non-normality or errors, individuals with progression scores 2.5 standard deviations from mean in 50% or more of the measures were removed from analysis. Additionally, all scores were adjusted using Box-Cox power transform to conform to regressions assumption of normally distributed residuals.[54]

Association and Meta-Analysis

Patients were split into two groups according to European or African American ancestry based on the genetic ancestry information, and analyses were pursued within each group. Using the additive allele effect-based regression analysis model implemented in PLINK, we tested single SNP associations against disease progression of GHVF sectors (i.e., slopes computed from the regression of visual loss on time) and summary global outcome measures.[55] SNPs were represented numerically as the number of minor alleles possessed by a patient at each locus: 0=homozygous wild-type or major allele, 1=heterozygous genotype; and 2=homozygous minor alleles. Covariates were also considered in the analyses and included age, time in the study, sex, and the first three principal components obtained from a Genetic Correlation Matrix (GRM) using PLINK for stratification and ancestry analysis. Results of the association analyses for each SNP by each GHVF phenotype were aggregated within genes (EUR N = 635, AAM N = 821) and intergenic regions defined by ANNOVAR variant notation.

For each of the fourteen phenotypes, aggregated association statistics were tested for enrichment of statistically significant ($p<5.0E-5$) associated variants using the COMBAT package , which provides a single p-value for a gene or region based on the strength of the association of each variant considered after correcting for linkage disequilibrium (LD) between the variants.[56] Ancestry matched genotype frequencies for the variants considered in our analyses, as well as insights into the LD structure among these variants, were obtained from the 1000 Genomes project data.[57]

Results

GWAS was performed with fourteen quantitative phenotypes, and we looked at enrichment within genes and intergenic regions across all phenotypes. Given the sample size, this study focused on gene level summary statistics via the use of the methods implemented in the COMBAT software suite rather than individual SNPs. We also examined loci already associated with POAG to see if they also exhibited signs of association with rate of disease progression.

We used SNPs and genes found to be associated with glaucoma in a recent publication (A multiethnic genome-wide association study of primary open-angle glaucoma identifies novel risk loci) for replication purposes as the authors of this publication used similar ancestry cohorts.[58] Replication tests in known SNPs for glaucoma onset were not ultimately conclusive (See supplementary material for greater details about these analyses).

For our analysis of patients with European ancestry, four genes previously associated with glaucoma diagnosis (TLR4, CDKN2B-AS1, ARHGEF12, TMTC2) were found to be associated with glaucoma progression, based on the gene-based test implemented in the COMBAT package, with progression defined as the slope of VF loss over time. In addition, a gene associated with a related eye disease (pseudoexfoliation syndrome), but not previously tied to glaucoma diagnosis, was also strongly associated with glaucoma progression (CNTNAP2 P=3.85E-07).[59]

In the African ancestry population several genes previously associated with glaucoma diagnosis were also associated with glaucoma progression (TMTC2, TGFBR3, CDKN2B-AS1). CNTNAP2 (P=1.67E-06) was strongly associated with progression in the African population as well. Table 1.2 provides a summary of the gene-based analysis results and Figures 1.2A and 1.2B provide a heatmap characterization of the genes associated with the different phenotypes for individuals of European and African American ancestry, respectively.

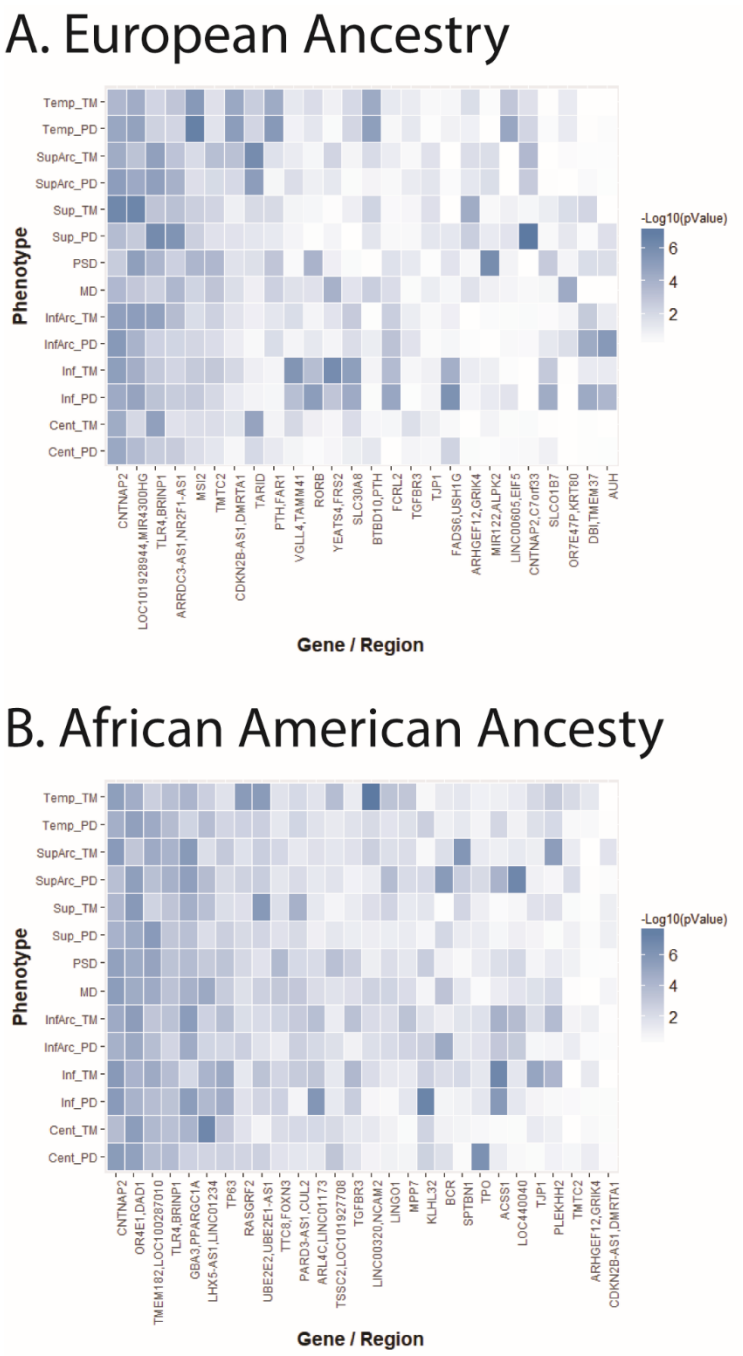


Figure 1.2 Gene by Phenotype Significance Enrichment
Shows significance by gene and phenotype arranged by average lowest p-value. Genes further on the left are more enriched for significance across all phenotypes. Both European ancestry and African American ancestry show CNTNAP2 as the most consistently significant.

Table 1.2 Associate Genes by Ancestry
Known genes to be associated with glaucoma onset which were also associated with progression in our study. Additionally, significant SNP's in Promoter / Enhancer regions that were identified with encodes RAMPAGE pipeline.

Population: European

Region	Chr	Previous Glaucoma Association	Gene P-Value	Signigicant SNP's in Promotor/Enchancer Regions
CNTNAP2	7	No	3.85E-07	6
CNTNAP2, C7orf33	7	No	6.41E-08	0
CDKN2B-AS1, DMRTA1	9	Yes	6.33E-06	0
TLR4, BRINP1	9	Yes	6.43E-07	0
ARHGEF12, GRIK4	11	Yes	4.42E-05	0
TMTC2	12	Yes	1.87E-04	0

Population: African American

Region	Chr	Previous Glaucoma Association	Gene P-Value	Signigicant SNP's in Promotor/Enchancer Regions
TGFBR3	1	Yes	9.44E-05	0
CNTNAP2	7	No	1.53E-06	2
TLR4, BRINP1	9	Yes	3.67E-05	0
TJP1	15	Yes	7.26E-06	0

Discussion

Glaucoma is a complex, multifactorial eye disease that has the largest prevalence of all neurological diseases. Identifying each and every factor contributing to glaucoma susceptibility will be difficult, given that each contributing factor may have a small effect. This is also true since glaucoma-related phenotypes, such as disease progression as characterized by VF loss, exhibit wide variation among patients, suggesting the existence of a pronounced cumulative or temporal component to the disease. We pursued a GWAS focused on glaucoma progression with a phenotype derived from measures obtained over an average of 9.1 years on 754 patients. We found that many genes that had been previously implicated in glaucoma susceptibility were also

associated with glaucoma progression, most notably, the gene CNTNAP2 and variants surrounding the CNTNAP2 gene. Interestingly, CNTNAP2 is expressed in all ocular tissues including glial and neuronal cells. Additional laboratory assays, constructs, and functional testing are required to validate our findings and put them into a more rigorous pathogenic context. We did cross-reference all SNPs that were significant at genome-wide levels with the Encode Screen tool[60,61] but found that there was no overlap between suggestive SNPs and promotor or enhancer positions shown to be functional. For CNTNAP2 we found six SNPs that were in five different enhancer or promotor regions (EH37E0933479, EH37E0933483, EH37E0933484, EH37E0933487, EH37E0933492). The RAMPAGE (RNA Annotation and Mapping of Promoters for the Analysis of Gene Expression) pipeline has identified all five of these regions as transcription start sites for CNTNAP2.

Our analysis was constructed to test several phenotypes efficiently. There are other ways we could have tested the effect of genetic variants within our population. As mentioned earlier, a report by Trikha et al.[37] described the results of an analysis of a binary 5-year 'progression or not' phenotype rather than a quantitative measure of progression. This has the advantage of not being heavily influenced by extreme cases but is probably ignores important variation in actual progression rates. Our association analysis considered the slope of an individual's worst eye as the dependent variable for the regression at each SNP. A more rigorous and computationally expensive analysis would use the raw VF measures as a dependent variable and implore more complicated regression techniques to handle serial correlation within each individual.

Our study is not without limitations, the first of which is the sample population and size. Our study was observational and thus limited by patient follow up and recruitment. While the study contained 1,219 individuals, 465 (38%) had to be removed for insufficient, incomplete, or questionable data. We have confidence in our findings, however, since our study involved a relatively small cohort, our main findings were observed in analyses involving two separate cohorts with different ancestral backgrounds, and therefore provides some assurance that they are true positives. In addition, our study leveraged WGS data, but only considered common variation in our GWAS. We did this in part because of power issues, but also because we did not have access to a sizable replication data sets to verify associations with rare variants. However, we do intend to assess the contribution of rare variants to glaucoma progression in the future.

Chapter 1, in full, has been submitted for publication of the material as it may appear in Ophthalmology Glaucoma, 2020, Argus J Athanas, Radha Ayyagari, Linda Zangwill, Mark Christopher, Jeffrey M. Liebmann, Christopher A. Girkin, Robert M. Feldman, Harvey Dubiner, Yii-Der Chen, Kent D. Taylor, Xiuqing Guo, Jerome I. Rotter, Nicholas Schork, Robert N. Weinreb for the ADAGES III Genomics Study Group. The dissertation author was the primary investigator and author of this paper.

CHAPTER 2 : LONG-TERM USE OF A MINDFULNESS AND MEDITATION SMARTPHONE APP IS ASSOCIATED WITH IMPROVEMENT IN BASELINE MOOD

Abstract

Background: The use of smartphone apps to monitor and deliver healthcare guidance and interventions has received considerable attention recently, particularly with regard to behavioral disorders, stress relief, negative emotional state, and poor mood in general. Unfortunately, there is little research investigating the long-term and repeated effects of apps meant to impact mood and emotional state.

Objective: We aimed to investigate the effects of both immediate point-of-intervention and long-term use (ie, at least 10 engagements) of a guided meditation and mindfulness smartphone app on users' emotional states. Data were collected from users of a mobile phone app developed by the company Stop, Breathe & Think (SBT) for achieving emotional wellness. To explore the long-term effects, we assessed changes in the users' basal emotional state before they completed an activity (eg, a guided meditation). We also assessed the immediate effects of the app on users' emotional states from preactivity to postactivity.

Methods: The SBT app collects information on the emotional state of the user before and after engagement in one or several mediation and mindfulness activities. These activities are recommended and provided by the app based on user input. We considered data on over 120,000 users of the app who collectively engaged in over 5.5 million sessions with the app during an approximate 2-year period. We focused our analysis on users who had at least 10 engagements with the app over an average of 6 months. We explored the changes in the emotional well-being of individuals with different emotional states at the time of their initial engagement with the app using mixed-effects models. In the process, we compared 2

different methods of classifying emotional states: (1) an expert-defined a priori mood classification and (2) an empirically driven cluster-based classification.

Results: We found that among long-term users of the app, there was an association between the length of use and a positive change in basal emotional state (4% positive mood increase on a 2-point scale every 10 sessions). We also found that individuals who were anxious or depressed tended to have a favorable long-term emotional transition (eg, from a sad emotional state to a happier emotional state) after using the app for an extended period (the odds ratio for achieving a positive emotional state was 3.2 and 6.2 for anxious and depressed individuals, respectively, compared with users with fewer sessions).

Conclusions: Our analyses provide evidence for an association between both immediate and long-term use of an app providing guided meditations and improvements in the emotional state.

Keywords

mental health; smartphone; emotional well-being; mindfulness

Introduction

Background

Behavioral conditions, neuropsychiatric diseases, and poor general mental health are seen as major contributors to morbidity, mortality, and lost productivity on a global scale. However, these factors are often overlooked in discussions about the current state of healthcare, which tend to focus on physical well-being.[44] Many studies suggest that mental health can play a large role in physical health, recovery from disease, and ultimately productivity and, therefore, should receive greater attention.[41,42,43] Unfortunately, there are serious questions about how mental health can be promoted and, in instances when it is called for, how relevant interventions can be prescribed and deployed efficiently in a cost-effective manner.[62,63,64] This is especially true given the number of people who may actually benefit from such interventions.[65] In light

of this, there is enthusiasm for the development of smartphone apps that can not only monitor an individual's health—both physical and mental—but also deliver content designed to help coach them through difficult times or provide a needed intervention. In fact, many smartphone apps have been developed, or are under development, to aid in healthcare via, for example, image-based diagnostics, glucose monitoring for diabetes, and physical fitness promotion.[66,67] For mental health management and intervention, there is growing enthusiasm for the development of smartphone platforms that provide guidance on mindfulness and meditation as a way of relieving stress and promoting mental health and well-being. Many of the resulting platforms have been or are undergoing testing in clinical studies.[68,69,70,71,72]

The use of mobile phone apps in combating or mediating behavioral conditions, stress, negative emotional states, and elevating mood is also consistent with directions that public health and regulatory officials are considering. In fact, evidence is mounting from clinical trials showing that smartphone apps can be effective in a variety of settings. Agencies such as the US Food and Drug Administration (FDA) have created, and in instances passed, legislation allowing the filing and approval of mobile health apps as approved health technologies on the same level as in vitro diagnostics and drugs. Pear Therapeutics was one of the first companies to have a smartphone app for addiction approved for use by the FDA in 2016.[73] Many other commercial and academic groups are developing smartphone apps for a wide variety of conditions that go beyond the simple direct-to-consumer market by seeking regulatory approval for their use in clinical contexts.[74,75,76] Unfortunately, not enough time has elapsed since the introduction of smartphone-based intervention apps to provide insight into their long-term repeated effects as well as their effects in real-world settings (ie, outside of clinical trials).[77,78,79]

Objectives

Stop, Breathe & Think (SBT) has developed a smartphone app that provides guided meditations and mindfulness activities to promote self-awareness coaching to interested users. As noted, mindfulness and meditation have been shown to improve affect and mood and promote healthy thought patterns.[45,46] The

SBT app prompts users before and after they are guided through meditation and mindfulness activities to provide an emotional, mental, and physical check-in, thereby allowing an assessment of an individual user's emotional state and mood pre- versus postactivity in real time. As repeated uses of the app by SBT users are archived, longitudinal information on its users with regard to their long-term engagement with the app is retained. This allows further analysis of the influence of repeated engagements with the app on an individual user's basal mood over time in real-world settings. We pursued such an analysis using data from SBT users who had at least 10 engagements with the app. The SBT app allows users to choose from more than 100 unique emotions to reflect their emotional state at the time they use the app. These emotions cover a range of human emotions including anger, remorse, anxiety, calmness, and enthusiasm. Users are guided through meditations that they can choose from based on an algorithm developed by SBT. We focused our analyses on the baseline (or basal) emotional state of a user, before he or she engaged in a guided meditation or mindfulness activity and were primarily interested in the long-term and repeated use effects of the SBT app on this baseline emotional state. Essentially, we wanted to ask the question if the continued use of the app lifted the spirits of the user over time. We were particularly interested in users who tended to pick emotions associated with depression and anxiety when engaging with the app before meditating.

Methods

The Basic Stop, Breathe & Think App

The SBT app is a multiplatform (ie, iOS, Android, and Alexa) app designed to guide users through meditations and mindfulness activities to alleviate stress, anxiety, and depression and improve the sense of well-being. Upon opening the app, a user can participate in an optional 10-second reflection period. After this optional reflection period, users describe their current mood, emotional state, and physical health by choosing from a number of emotions; the SBT app then provides suggestions for specific meditation and mindfulness activities. The user can choose from among the suggested activities after being asked to endorse up to 5 different characterizations of their mood and emotional state. A user can choose not to provide any input regarding their mood, emotional state, and physical health and simply engage in an

activity. The figure below (Figure 2.1) provides a schematic of an individual session and the corresponding points where user information is collected.

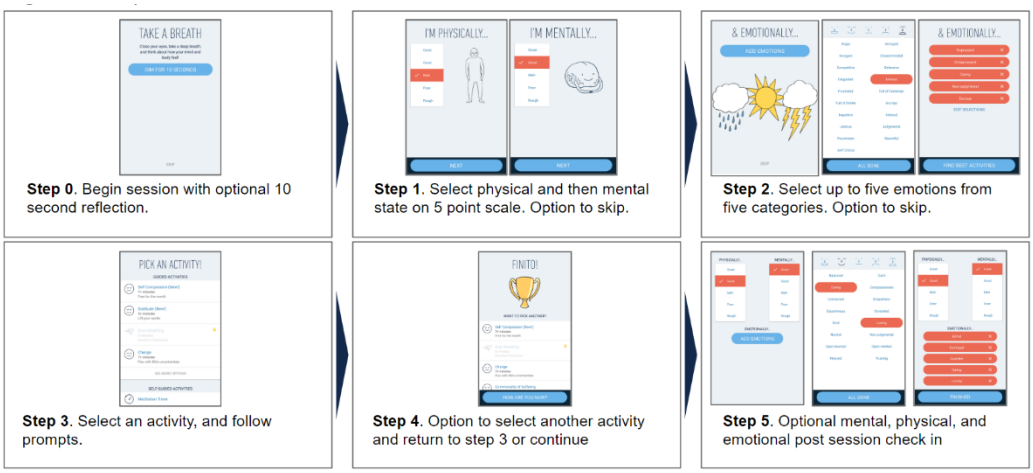


Figure 2.1 Stop Breathe Think User Interaction Flow Chart
Stop, Breathe & Think user interface and stages of interaction with the app. Users are provided several ways in which they can record their current emotional state both pre- and postactivity. These emotional check-ins are optional, but the intuitive and simple selection process makes it easy for most users to enter at least some emotional status information.

It should be understood that all information collected with the SBT app is volunteered by users as stated and defined in the SBT user licensing agreement and privacy policy. In addition, for purposes of our data analyses, all the data we obtained from SBT were anonymized and put into a Health Insurance Portability and Accountability Act (HIPAA)-compliant format such that users could not be reidentified. Functionality and delivery of the SBT app and service varies from device and platform implementation (eg, Alexa, Android, and Web browser). Therefore, to avoid batch effects, we focused on users who were exclusively on an iOS platform and started using the app after SBT provided its last major version of the app (05/01/2016). Users had to have completed at least 10 sessions or engagements with the app, with a minimum of 6 of those sessions including pre- and postactivity emotion selections. The SBT app content is in English and to avoid translation errors and alternative interpretations of the language used in the SBT app, we restricted our analyses to individuals from native English-speaking countries: The United States, United Kingdom, Canada, and Australia. An additional filter was used, restricting users' ages to between 12 and 100 years.

Emotional Check-ins Pre- and Postactivity Score

The SBT app allows the user to endorse between 1 and 5 emotional states out of a possible 115, before and after engagement in a guided meditation or mindfulness activity (or series of activities if they choose to engage in more than 1 activity during a session). This emotional check-in involves selecting an initial emoticon and then choosing from a list of emotions within subgroups of terms that closest characterize the user's current emotional state. These 115 emotions were chosen for the app based on internal SBT research and user requests. All emotions were classified as positive, neutral, or negative and given corresponding scores of 1, 0, and −1, respectively. All emotions and their corresponding scores are provided in Multimedia Appendix 1. As users can select up to 5 emotions, an average emotional score was calculated for both pre- and postactivity and standardized to a range from −1 (all negative emotions) to 1 (all positive emotions). Our analysis explored (1) trends in the preactivity emotional score over repeated uses of the app while accounting for the covariates as well as serial correlation between sessions and (2) trends in changes of the emotional scores before and after an activity over repeated uses of the app.

Clustering of Emotions

In addition to treating the preactivity emotion scores and changes in emotion scores pre- and postactivity as dependent variables and time, sex, and age covariates as independent variables, we also explored the patterns among the emotion endorsements to see if there was evidence for obvious clusters of emotions that could reflect the same general emotional state. We leveraged principal coordinates analysis (PCoA) and the nonsupervised clustering technique, Partitioning Around Medoids (PAM), for these analyses.[80] We pursued these analyses as it is arguable that some users may see a subset of the emotions as synonymous and hence only choose one among many possible choices to describe their emotional state at the time to avoid redundancy, whereas other users might see those same subsets of emotions as complementary and reflecting different aspects of their mood. In addition, other users may preferentially

select emotions based on their location in the selection list or choose a set of rare emotions that are infrequently selected by other users to differentiate their emotions.

The distance between the emotions was calculated using the Bray-Curtis distance measure.[81] To determine the optimal number of non-supervised emotion clusters in 2-dimensional PCoA component space, we selected the number of clusters with the largest silhouette score. Once we identified the optimal number of clusters, emotions were then assigned to one of the identified clusters.

An individual's emotional status was also summarized in terms of the relative distances (using the Euclidean distance measure) between pre- and postactivity states. The distances between an individual's emotional status and the medoid of the closest associated emotion cluster were calculated as well. Emotions were labeled with clinical categories, associating each of them with either anxiety, depression, anger, or happiness (Multimedia Appendix 1). Ultimately, using distances between emotional states and emotional clusters allowed us to build models relating the number of times users engaged with the app to gross changes in emotional states defined by the emotion clusters.

Statistical Analyses to Identify Long-Term Changes in Emotional State

To assess the effects of the continued use of the app on the preactivity emotional state, we used Linear Mixed-Effects (LME) models and Generalized Linear Models (GLMs) as implemented in the lme4 package in R.[82] These analysis techniques can accommodate serial correlations among emotions over time and also account for both fixed (eg, sex) and random effects (eg, variation in preactivity emotional state or the degree to which use of the app changed the preactivity emotional state over time). We pursued different analyses to evaluate changes in the preactivity emotional state over time, including a model that considered the effect of the emotional states possessed by individuals at their first engagement with the app. These analyses considered both the emotion scores as the dependent variables as well as the use of the emotions as defined by the cluster analysis clinical labels as dependent variables. We also tested the effect of repeated

uses of the app on the change in the emotional state pre- to postactivity by treating the ratio of pre- to postemotion score as a dependent variable.

We included several covariates in our analyses and tested them for their effects on the emotional state: session index (ie, 1 as the first use and 2 as the second use—which captures the repeated use of the app), gender, age, country of origin, subscription status, and whether the user remained anonymous (ie, did not fill out information in his or her account—which may indicate a fake or disengaged user). As there is large variability in the number of completed sessions and the distribution of the number of uses of the app per individual has an extreme right skew, we applied a log10 transformation to the session index variable. This transformation markedly improved the normality of the session index as a variable (data not shown). LME models were fit, and the features associated with the preactivity emotional state as the dependent variable were selected using a forward stepwise selection procedure based on the Akaike Information Criteria. Similar models were fit with the pre- to postactivity emotional state ratio as the dependent variable. GLMs were fit to the data when changes in emotion categories (ie, based on clinical or cluster analysis labels) were taken as the dependent variable.

Results

Defining the Dataset

After all the duration, quality, platform, and country filters were applied, 13,393 users remained (10,082 females, 2187 males, and 1124 undeclared sex). The average age of the users was 32.3 (SD 13.5) years, with 31.7 (SD 13.3) years for females, 34.6 (SD 13.4) years for males, and 33.3 (SD 15.0) years for undeclared participants. Collectively, the users completed 569,961 sessions with the app, with 302,514 of these sessions having emotional check-in data, with an average of 42.6 sessions and 22.6 emotional check-ins per user. Multimedia Appendix 2 provides a histogram depicting the distribution of the length of time users engaged with the app. Multimedia Appendix 3 shows average period between app uses given the total length of engagement for users.

Cluster Analysis of the Emotions

The use of the silhouette scores based on the PCoA and PAM analyses suggested that there were likely 8 clusters of emotions.[83] As noted, the relative distances between pre- and postactivity emotional states and the distances between each user's emotional state and the closest associated emotion cluster were calculated. In addition, each of the 115 emotions that could be endorsed was assigned to one of the emotion clusters (see Multimedia Appendix 1). Using these cluster labels, we calculated the mean orientation of each cluster and the relative distance of each individual's emotional scores both pre- and postactivity from these means. These distances were compared with the other emotion scores we calculated and were highly correlated with them (Figure 2.2). Figure 2.3 provides a graphical depiction of the results of the clustering using the first 2 principal coordinates obtained from our analyses.

Figure 2.2 Variate Correlation

Average emotional score versus cluster centroid distances correlation matrix represented as a heat map. As an example for interpreting the numbers in the matrix, a −0.90 correlation between the preactivity emotion score (x-axis Average Pre Emo Score label) and positivity cluster (y-axis Dist positivity label) shows that users who score higher on the preactivity emotional score had a shorter distance of their selected emotions to the centroid of the positive emotion cluster. Note that labels with Dist reflect distance measures derived from the cluster analyses (eg, Dist Anxiety reflects the distance of a user's emotional score from the anxiety cluster mean) and Emo reflects a specified emotional cluster.

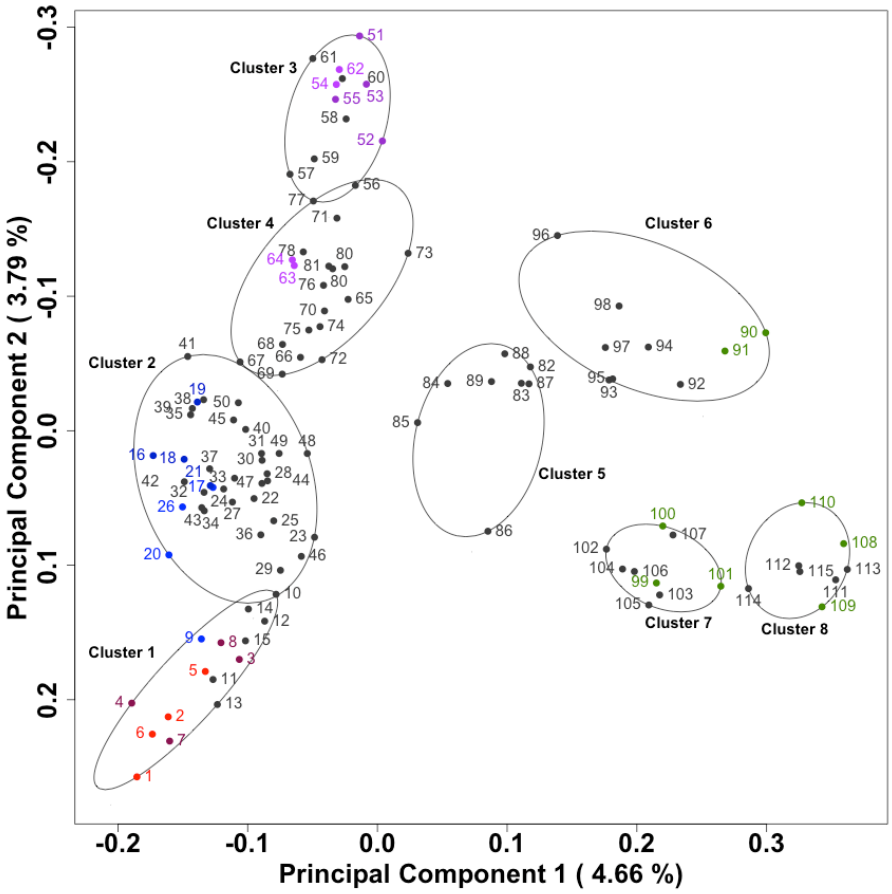


Figure 2.3 2-Dimensional PCoA Clustering Based on Co-selection
Emotion clustering using both pre- and postactivity emotion endorsements. The points in the plot reflect positions in the first 2 principal components defined by the Bray-Curtis distance between each pre- and postactivity emotional selection. The 8 circular clusters encompassing the emotions were defined by a permutation around medoids analysis technique, in which 8 clusters maximized the average cluster silhouette scores. Cluster boundaries are drawn on the smallest region including all underlying emotions. Emotions are labeled by clinical association such that terms clinically associated with anger are in red and pink, depression in blue, anxiety in purple, and happiness in green.

Table 2.1 Assignment and scores for Stop, Breathe & Think selectable emotions.
All endorsable emotions within the SBT application with clustering, clinical assignment, and emotional score. Not all emotions were given clinical assignments.

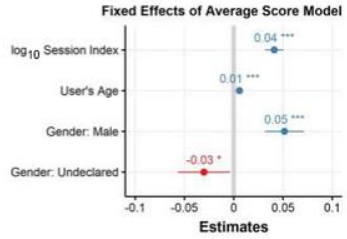
Assigned Cluster	Clinical Assignment	Emotion	Assigned Emotional Score
1: Anger	Angry	Angry[1], Annoyed[2], Disgusted[3], Frustrated[4], Grumpy[5], Irritated[6], Resentful[7]	-1
		Impatient[8]	0
	Depressed	Pessimistic[9]	-1
	None	Closed-minded[10], Defensive[11], Full of Contempt[12], Full of Dislike[13], Jealous[14], Judgmental[15]	-1
2: Depression	Depressed	Depressed[16], Disappointed[17], Discouraged[18], Sad[19], Self Critical[20], Unhappy[21]	-1
	None	Alienated[22], Arrogant[23], Ashamed[24], Bullied[25], Defeated[26], Despairing[27], Despondent[28], Envious[29], Grieving[30], Guilty[31], Heartbroken[32], Helpless[33], Hurt[34], Insecure[35], Insulted[36], Isolated[37], Lacking Confidence[38], Lonely[39], Longing[40], Overwhelmed[41], Powerless[42], Rejected[43], Remorseful[44], Vulnerable[45]	-1
		Competitive[46], Embarrassed[47], Homesick[48], Not Confident[49], Sensitive[50]	0
3: Anxious	Anxiety	Anxious[51], Concerned[52], Nervous[53], Restless[54], Worried[55]	-1
	None	Apprehensive[56], Disconnected[57], Distracted[58], Mixed Up[59], Scattered[60], Stressed[61], Uneasy[62]	-1
4: Tired	Anxiety	Afraid[63], Panicked[64]	-1
	None	Achy[65], Clingy[66], Exhausted[67], Fatigue[68], In Pain[69], Insomnia[70], Lazy[71], Suspicious[72], Tired[73], Torn[74]	-1
		Apathetic[75], Bored[76], Conflicted[77], Confused[78], Guarded[79], Hesitant[80], Indifferent[81]	0
5: Compassion	None	Caring[82], Compassionate[83], Empathetic[84], Equanimous[85], Fiery[86], Kind[87], Non-judgmental[88], Trusting[89]	1
6: Calm	Positive	Calm[90], Relaxed[91]	1
	None	Balanced[92], Connected[93], Grounded[93], Loving[95], Neutral[96], Open-hearted[97], Open-minded[98]	1
7: Excited	Positive	Enthusiastic[99], Excited[100], Satisfied[101]	1
	None	Creative[102], Fulfilled[103], Glad[104], Joyful[105], Proud[106], Relieved[107]	1
8: Grateful	Positive	Content[108], Happy[109], Hopeful[110]	1
	None	Appreciative[111], Encouraged[112], Grateful[113], Strong[114], Thankful[115]	1

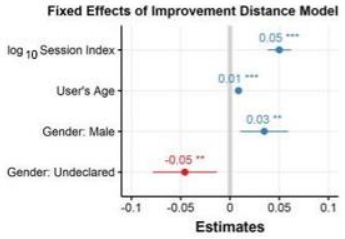
Mixed-Effects Modeling: Long-Term Use Effect on Preactivity Mood and Emotional State

Using the average preactivity emotional scores, as well as the cluster-based distance measures, as dependent variables, we fit linear-mixed models with session, as well as the important covariates, as independent variables, while accommodating serial correlation emotions. The results using the average preactivity emotional state scores suggest that a statistically significant relationship exists between the number of uses of the app (ie, session index) and the preactivity emotional state, with an elevation in mood (i.e, increase in positive emotions) occurring with repeated use of the app. Adjusting for scale, users experience a 2% improvement in mood after their first session, a 4% increase after their 10th session, and a 6% increase after their 100th session. The clinical relevance of this improvement in mood needs to be investigated further. We found that males have an average 2.5% higher (improved) preactivity mood than females and that older users have a more positive mood than younger users. Additional analyses suggested that repeated use of the app resulted in specific improvements in levels of anxiety and depression. After the first 10 sessions with the app—which on average corresponded to a 63.4-day period—users were 82% more likely to report no anxious emotions and 28% more likely to report no depressive emotions. This effect was even more pronounced when we only examined users whose first emotion endorsement reflected anxiety (440%) or depression (1050%). Figure 2.4 depicts the effect size and statistical significance of the estimated regression coefficients for the analysis models with the average emotional score in the left panels and cluster-based emotion similarity scores in the right panels. The statistical significance (ie, P values) were calculated using a Wald-Z statistic approximation. Models fit using a subset of users who reported anxious or depressed emotions in their first session with the app are labeled as primary models. The session index is consistently associated with improvements in mood, suggesting, again, that repeated use of the app positively impacts mood.

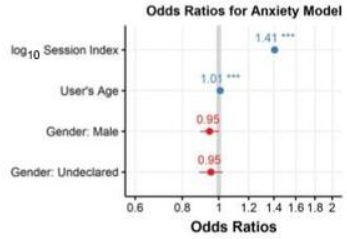
Figure 2.4 Basal Emotional Model Estimates
Linear mixed-effects regression coefficient estimates, their SEs, and P values (<.001***, <.01**, and <.05*) for models with the preactivity emotional state as the dependent variable. Analyses with the emotion scoring method as the dependent variable are on the left panels and analyses using distances from clustering as the dependent variable are on the right panels. Generalized Linear Model logit regression models were used with a binary dependent variable indicating if the emotion terms endorsed at a session reflected anxiety (middle panels) or reflected depression (bottom panels)

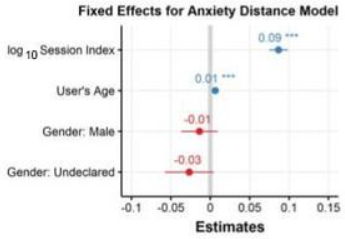
Basal Overall Mood Improvement

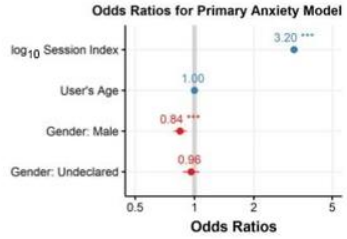


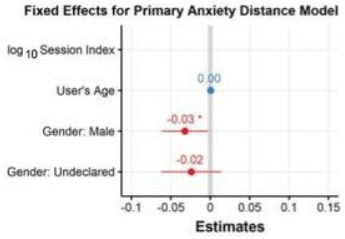


Basal Changes in Anxiety

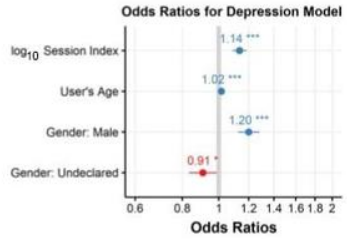


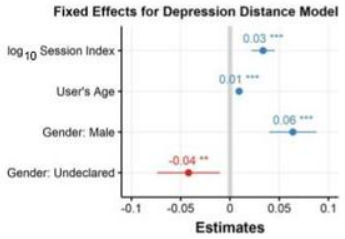


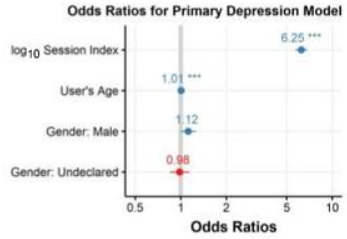


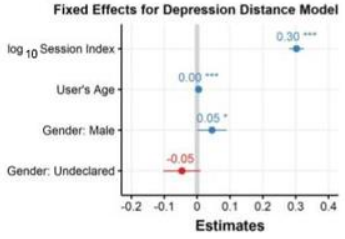


Basal Changes in Depression









Mixed-Effects Modeling: Pre- Versus Postactivity Mood or Emotional State

We also fit models that considered the ratio of preactivity to postactivity emotional scores as the dependent variable. Figure 2.5 plots the regression coefficients resulting from the fits of these models with the ratio of average emotional score pre- to postactivity as the dependent variable (top panel) and the ratio of the distances between the emotions based on the clustering (bottom panel). The results suggest that repeated use of the app leads to increases in improvement of the mood/emotional state achieved through a meditation or mindfulness activity—or rather that the activities seem to lead to larger improvements in mood as the user has more engagements with the app.

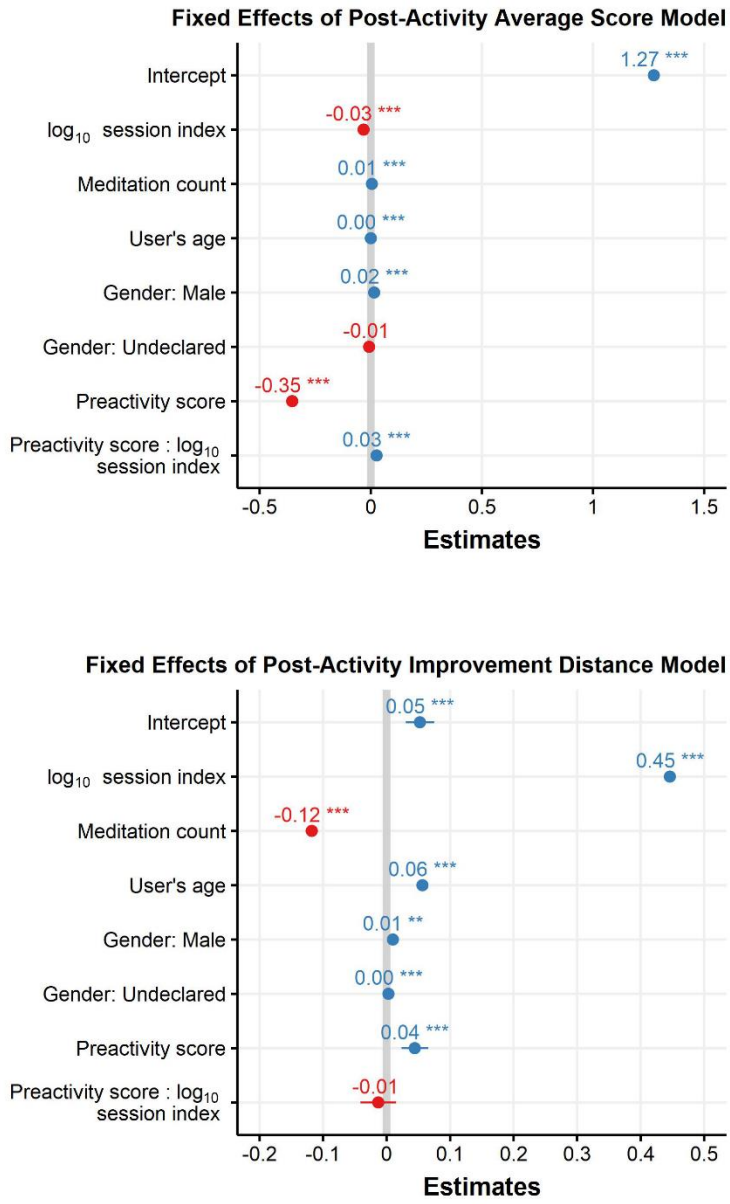


Figure 2.5 Immediate Emotional Change Model Estimates
Linear mixed-effects regression coefficient estimates, their SEs and P values (<.001***, <.01**, and <.05*) for models with pre- to postactivity change in the emotional state as the dependent variable. An analysis with the standardized change in emotion score pre- to postactivity as the dependent variable is reflected in the top panel, and proximity to the positive emotional clusters as the dependent variable is reflected in the bottom panel.

Discussion

Principal Findings

Our analyses show that repeated engagements with the SBT app are associated with an improvement in users' emotional states over time. In the absence of a randomized control trial, it is difficult to say with certainty that there is a direct causal relationship between the use of the SBT app and emotional state; however, given the large diverse sample size, we believe that the impact of unmeasured covariates on our results (such as external events in the users' lives) is likely to be small, although potential biases in the users of the app may exist. The effect we observed is more pronounced for users who often endorse anxiety or depression when capturing their emotional state at their initial uses. We also found that age and sex covariates are associated with the basal mood or emotional state. Ultimately, our analyses suggest the possibility that guided meditations and mindfulness activities have the potential to be effective ways of reducing anxiety, depression, and stress and ultimately elevating mood, although the ultimate clinical significance of the improvements in the emotional state that we observed needs to be explored. Our analyses did reveal other interesting phenomena. For example, although a minority in our study, males tended to have higher baseline emotional scores and responded better to the SBT app than females. The age of a user was also found to be a significant correlate of the basal emotional state, with older users generally endorsing more positive emotions.

Limitations of the Study

Our analyses are not without limitations, the first and foremost being that there is no control group and comparator app. This makes it difficult to definitively state that guided meditation and mindfulness activities are causally related or responsible for the increase in baseline mood or emotional state over time. However, given the sample size and magnitude of the effect, the significant change in emotional state after immediate and prolonged use of the app suggests that it has potential as an intervention. Another limitation is that all the information we analyzed was self-reported without any oversight by a third party. There could be users who did not follow instructions and entered erroneous emotions to expedite engagement with the

meditations. Many of the individuals we did include in our analyses did not record emotions for each and every one of their sessions, resulting in many incomplete observations. Finally, a potential limitation with our analyses is that there could have been a heavy selection bias among the individuals using the app in the sense that they were motivated enough to download it and use it. Thus, this may be an indication that they could be predisposed to responding positively to the app.

Broad Emotional State Transitions

Our use of the emotion clusters and similarity scoring of emotions based on our cluster analyses of those emotions allowed us to explore how often individual users transitioned from one broad set of analogous and almost synonymous emotions to another. On the basis of these analyses, we found evidence that, in general, individual users' emotional states move from negative to positive over repeated uses of the app. We find that anxiety-prone and more depressed individuals benefit from the app more than others. These findings, as with the analyses, need to be verified in more controlled settings, such as randomized control trials, but again suggest that there is promise for the app and related apps in clinical and public health settings.

Future Directions

There are a number of questions that deserve attention beyond those that we addressed with our data. For example, the number of uses of the app may not reflect the total length of time the app was used (eg, a user could engage with the app intensely over a short period of time or stretch their use out over a longer period of time). Assessing the impact of the number of uses versus length of time on outcomes could provide a more detailed insight into the benefits of the app. In addition, it would be good to see if a companion study designed especially for adolescent populations also has a positive effect on their emotions.[84] In addition, special clinical populations may benefit from the app (eg, clinically depressed individuals and individuals with addictions). It would be of value to explore analyses that focus on the impact of large-scale social stressors (eg, school shootings, national election results, and natural disasters)

on the use of the app as well as its effects on mood in the wake of stress-inducing events. Geolocation data on users could better define such exposures to social stressors should they be location specific (eg, a natural disaster in a particular state). Finally, as emphasized, it would be ideal to test the utility of the app in bona fide clinical trials to determine which aspects of the app are causally related to improvements in mood and emotional state as well as identifying subgroups of individuals that appear to respond best to particular activities.

As more and more attention is given to the delivery of healthcare and health maintenance strategies through devices such as smartphones, robots, and telemedicine communications, greater sensitivity to the nuanced effects of these devices should motivate studies of them that are pursued in a comprehensive manner. Such sensitivity and more elaborate studies could also lead to more efficient and sophisticated deployment of these devices and help combat the need for expensive and logistically challenging visits to healthcare providers.

Conflicts of Interest

SS an advisory consultant and holds equity in Stop Breathe & Think.

JP is a cofounder of Stop Breathe & Think and holds equity in Stop Breathe & Think.

JG is an employee of Stop Breathe & Think and holds equity in Stop Breathe & Think.

JC is a cofounder of Stop Breathe & Think and holds equity in Stop Breathe & Think.

NJS is an advisory consultant and holds equity in Stop Breathe & Think.

Abbreviations

FDA: Food and Drug Administration
GLM: Generalized Linear Model
LME: Linear Mixed-Effects
PAM: Partitioning Around Medoids
PCoA: principal coordinates analysis
SBT: Stop, Breathe & Think

Chapter 2, in full, is a reprint of the material as it appears in Journal of Medical Internet Research, 2019, Argus J Athanas, Jamison M McCorrison, Susan Smalley, Jamie Price, Jim Grady, Julie Campistron, Nicholas J Schork. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3 : CHARACTERIZING EMOTIONAL STATE TRANSITIONS DURING PROLONGED USE OF A MINDFULNESS AND MEDITATION APP: AN OBSERVATIONAL STUDY

Abstract

Background: With the rising need for mental healthcare, shortages in mental healthcare providers, and unequal access to care, digital device apps that can be used to adjust mood show promise for helping meet mental healthcare demands. The therapeutic content delivered through a digital therapeutic to combat aspects of mental health disorder can also be 'personalized,' and hence potentially provide greater benefit.

Objective: We sought to characterize the transitions from one emotional state to another during the prolonged use of a digital app designed to provide a user with guided meditations based on their initial emotional state. Understanding the factors that mediate such transitions can lead to improvements in predictions and recommendations for which specific Mindfulness and Meditation Activities (MMAs) might be appropriate.

Methods: We analyzed data collected during the use of the Stop, Breathe and Think (SBT) mindfulness app. The SBT app prompts users to input their emotional state prior to and immediately after engaging with MMAs recommended by the app. Data were collected on more than 650 thousand SBT users involving over nearly 5 million MMAs. We limited the scope of our analysis to users with 10 or more MMA sessions and data on at least 6 basal emotional state evaluations. Using clustering techniques, we empirically grouped emotions into more coherent groups and then applied longitudinal mixed effect models to determine the effects that individual recommended MMAs had on emotional state transitions.

Results: We found that basal emotional states have a strong effect on transitions to a different emotional state after MMA engagements and that different MMA impact these transitions. We found that MMAs were effective in eliciting a healthy transition but only under certain conditions, and also observed gender and age effects on these transitions.

Conclusions: We find that SBT MMA app users' initial emotional state has an impact on which MMAs will have a favorable effect on their transition to another emotional state. Our results have implications for the design and use of guided recommendations for digital therapeutics.
Keywords: Mental Health, Mobile, Smartphone, Emotional, Mindfulness

Background

The call to treat mental health disorders on a footing equal to physical health disorders is growing.[85-87] Mental health disorders are known to affect productivity, comorbid conditions and physical wellbeing generally.[41,44,88] In fact, when asked, nearly ninety percent of Americans value mental health on par with that of physical health.[89] This is not without reason, as the prevalence of anxiety disorders alone is estimated to be between 3.8-25.0%.[65] Given the high collective prevalence of mental health disorders, their impact on quality of life, and the costs associated with the care of individuals with mental health disorders, there is a huge need to develop more efficient and reliable ways of not only treating them but also preventing them. Unfortunately, developing the appropriate infrastructure to combat mental health disorders within the current health systems will be daunting and expensive as many people find available mental healthcare overly complicated and often inaccessible.[90,91] Fortunately, newer and more accessible approaches to the care of individuals with mental health issues are being developed and include the use of telehealth, an emphasis on risk mitigation as opposed to treatments, and the use of digital therapeutics.[92] Of these, digital therapeutics are receiving a great deal of attention. Basically, digital therapeutics are digital devices (e.g., smartphone apps) that provide content meant to provide guidance on dealing with symptoms or content meant to alleviate symptoms in some other way (e.g., via imagery).

The emergence of digital therapeutics is recognized by public health and government regulatory agencies as well. The Food and Drug Administration (FDA) has allowed mobile mental health apps to receive approvals and accreditation as bona fide medical health apps just like drugs.[73,93-95] Care that includes the use of digital therapeutics can be scaled to help meet the demand without requiring many additional trained professionals. More importantly, digital therapeutics have great potential to help provide care in underserved populations where financial, professional scarcity, and societal burdens make other forms of care inaccessible.[96] Digital therapeutics have obvious limitations in that they aren't appropriate for use in all settings. One example of where they make good sense is in behavioral and mental health settings

involving stress management where techniques, such as encouraging relaxation via, e.g., meditations and mindfulness, can be used. In fact, mindfulness and meditation activities have been linked with healthy thought patterns and improved mood and could reduce stress and anxiety that is often of a type that is a precursor or symptom of many mental health disorders.[45,97]

Unfortunately, while the promise of digital therapeutics for reducing the risk and treating some mental health issues and disorders are great, there is a need to vet different digital therapeutics and understanding the settings in which they might be most effective. This is due in part to difficulties in defining mental health disorders and tracking individuals' symptoms over time in a way that can shed light on when to intervene and in what manner. This is as true for very serious mental health disorders, such as treatment resistant depression, as it is for managing stress and anxiety. For example, determining which personal settings and emotional states are appropriate for different interventions, such as Mindfulness and Meditation Activities (MMAs), have yet to be explored in full.[98] In fact, it is quite likely that there is a great deal of intra and inter individual variability in mood and feelings of stress and anxiety that might be necessary to understand and characterize so that guidelines and interventions, such as MMAs, can be tailored or personalized to individual users.

Objective

We have pursued a series of analyses to explore how the moods of the users of the SBT MMA app change or transition to other moods as they engage the app. As discussed in our prior paper on data collected via the SBT MMS app,[99] the SBT app recommends sets of MMAs to users based on their mood at the time they engage the app. In our initial work, we identified a trend for improvements in basal mood with prolonged use of the app. We also showed that, on average, a user's mood was improved after a single session with an MMA recommended by the app. In this paper, we aim to assess the specific effects that the recommended MMAs have on the transitions between moods of users at baseline and after they participate in a recommended MMA. A better understanding of which MMAs drive changes in emotional state could

lead to insights into which MMAs might be appropriate for individuals with specific emotional state profiles. Such an understanding could lead to better predictions and recommendations for individual users.

Methods

The App

The app developed by SBT is designed to guide users through MMAs which are created to reduce stress, anxiety, depression, and improve internal focus, mood, mental state, and sleep. This app is multiplatform and can deliver MMAs through many different platforms (i.e. iOS, Android, Alexa). When the app is engaged, a user is prompted to perform an optional 10-second reflection, which is followed by optional prompts to state how he or she is feeling mentally, physically, and emotionally. The mental and physical questions ask users to respond on a five-point scale with the following categories: great, good, meh, poor, rough. Following this, users are asked to pursue an 'emotional check-in' involving a selection of 1 to 5 terms from a pool of 115 emotions which describe their current state. After this initial check-in, users are shown some suggested MMAs (e.g., *Gratitude, Silence, Breathing*), but are free to select an MMA of their preference within a defined set. After the completion of an MMA users can continue to select additional MMAs or end the session. At the end of the session users are again prompted to do another optional check-in for mental, physical, and emotional state. A flow diagram of the user experience with the SBT app can be seen in Figure 3.1.

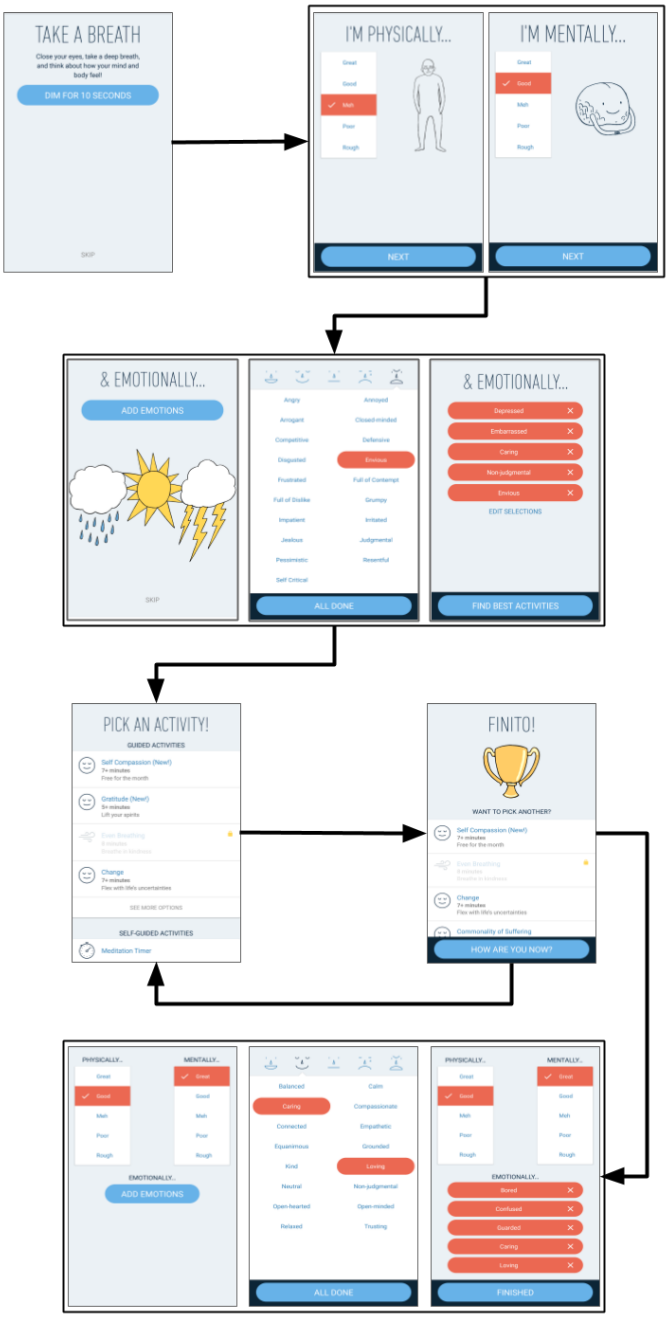


Figure 3.1 SBT App User Flow Diagram
A diagram depicting the users' experience when engaging the SBT app. A natural flow allows the user to reflect, check-in, perform an activity, and then check-in again. Reflections and check-ins are optional but were required data points for our analysis.

All data collected through the SBT app is volunteered by users as stated in the SBT user licensing agreement and privacy policy. All reported data were anonymized and put into a Health Insurance Portability and Accountability Act (HIPAA) compliant format so that users could not be reidentified.

Our data preparation methods were nearly identical to those described in our previous publication on the SBT data; however, in the current analysis data from several users were removed from all analyses by direction from the legal and compliance team at SBT. Because of our de-identification process, we cannot distinguish which users were removed from the previous analysis, or which users are now included because they have since met active user filtering criteria. We can say based on information about when users started to use the SBT app that there were 856 new active users who passed our filtering criteria since our last publication, and at least 3,219 users were removed in this analysis from our last publication. The SBT app has variation in functionality and delivery across platforms, and to avoid confounding effects we focused our analysis only on iOS users. We excluded users who joined before the last major update (05/01/2016). Further filtering was pursued to only include active users (10+ sessions completed) who had at least six sessions with both pre and post emotional check-ins filled out. To avoid cultural differences and language barriers, only users from English-speaking countries (United States, United Kingdom, Canada, and Australia) were included. Lastly, for both compliance and questions about the accuracy of the information we excluded users under the age of 13 or over the age of 100 from analysis.

Clustering Emotions

As described in our previous paper, emotions endorsed by users of the app were grouped into clusters based on the user's co-selection of emotional terms. Emotions were compared using Bray-Curtis dissimilarity.[81] We used Principal Coordinate Analysis (PCoA) to translate emotional dissimilarity into two-dimensional space using the first two PCoAs, and then used Partitioning Around Medoids (PAM) along with silhouette scores to determine the optimal number of clusters.[83] Each of the 115 emotions was then assigned to a cluster and the corresponding cluster medoid recorded. Individual emotional states, both pre

and post MMA engagement, were defined by the nearest (in terms of Euclidean Distance) cluster. These clusters define distinct emotional state categories, and the change (or 'transition') in emotional state categories between pre-MMA and post-MMA is the focus of our analysis.

To better understand and synthesize the results of our emotion clustering, we projected the clusters onto the Yale Mood Meter (YMM; Figure 3.2).[100] This YYM groups emotions into four quadrants which are defined by the 'energy' of the emotion (y -axis) and the 'pleasantness' of the emotion (x-axis). We color code these quadrants using the accepted criteria: Red = high energy, low pleasantness; Yellow = high energy, high pleasantness; Blue = low energy, low pleasantness; and Green = low energy, high pleasantness. We assigned each of the clusters to each of the four quadrants based on the majority of emotions within each cluster that mapped to a quadrant. This mapping allows us to think about transitions from quadrant to quadrant instead of simply cluster membership, which has several advantages: 1) it reduces our search space from 64 transitions to 16; 2) it increases the sample size for each transition thus providing better power to detect changes; and 3) it provides an interpretable scale for transitions (i.e., a high energy, low pleasantness state [red] to high energy, high pleasantness state [yellow]). While users may have different objectives in engaging the app and an MMA, the intuition is that red and blue states are undesirable, whereas yellow and green are desirable.

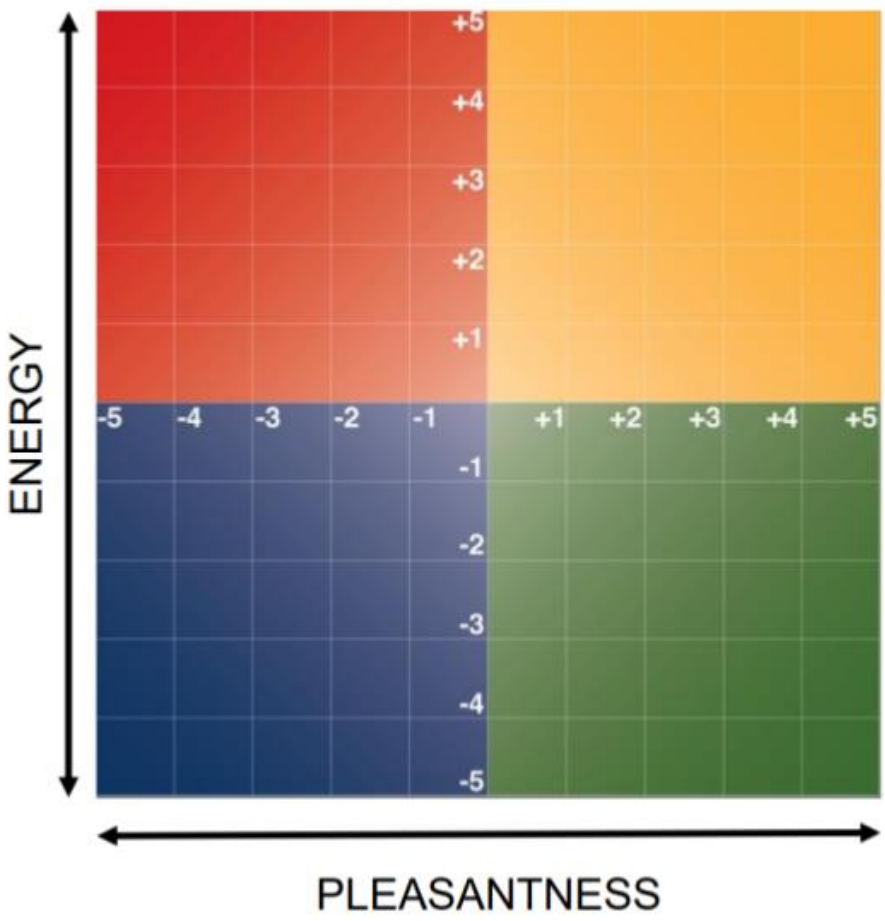


Figure 3.2 Yale Mood Meter
This framework was designed by Marc Brackett, PhD. classifies emotions into a 2-dimensional space with pleasantness as the x-axis and energy as the y-axis. The negative emotion quadrants red and blue represent low pleasantness, Red is higher energy such as anger, and blue is low energy such as sadness. The more favorable quadrants, green and yellow, are high pleasantness with energetic emotions like 'excited' fitting in yellow and lower energy emotions like calm fitting in green.

Description of MMAs

The SBT app provides over 100 different MMAs for users to choose from, with varying levels of popularity. Given the number of MMAs and the risk for overfitting our analysis models by considering each MMA individually, we focused on the top 20 chosen MMAs (86.8% of all completed MMAs) and combined the rest into a single category, 'Other', for a total of 21 MMAs assessed in analyses. The distribution of these MMAs and their descriptions can be found in Table 3.1.

Table 3.1 Meditation and Mindfulness Activity Frequency
Descriptions and frequencies of the top 20 MMAs completed. The remaining MMAs are aggregated as 'other'.

MMA Name	Count	Percent of all MMAs	Cumulative Percent of all MMAs	MMA Description
Relax, Ground and Clear	35,863	12.4%	12.4%	Use your imagination to relax, feel more stable and grounded, and clear your mind.
Reflection	32,528	11.2%	23.6%	Meditation timer with customizable chimes and soundscapes
Falling Asleep	31,153	10.8%	34.4%	Let go of the day's stresses and bring a sense of open, calm awareness into sleep.
Gratitude	24,744	8.6%	43.0%	Boost resilience and positive feelings by focusing on gratitude
Body Scan	19,943	6.9%	49.8%	Bring awareness to each part of your body, noticing your experience with a sense of curiosity and openness.
Kindness	12,368	4.3%	54.1%	Feel happier, more connected and strengthen your heart by focusing on Kindness
Mindful Breathing	12,094	4.2%	58.3%	Focus your attention on your breath, noticing what's going on inside (your thoughts, feelings, and physical sensations) and outside (your interactions and surroundings) with openness and curiosity.
Joy	10,691	3.7%	62.0%	Magnify your positive feelings by celebrating the positive moments, accomplishments, happiness, skills and good qualities of yourself and other people.
Noting	10,569	3.7%	65.6%	Strengthen the mind's capacity to stay present, so it's not always jumping from thought to thought, like a monkey jumping from tree to tree.
Welcoming the Day	8,280	2.9%	68.5%	Begin the day mindfully by focusing on your body's natural feelings of warmth and peace, and carry those warm feelings into your day.
Great Compassion	6,498	2.2%	70.8%	Break down the barriers between yourself and others by breathing in their difficulties and breathing out kindness and compassion.
Mindful Walk	6,036	2.1%	72.8%	Become more attentive to what's going on inside (your thoughts, feelings, and physical sensations) and outside (your interactions and surroundings) without getting caught up in judging or evaluating
Equanimity	5,817	2.0%	74.8%	Develop a sense of mental calmness and evenness towards all people and situations, and allow yourself to extend your kindness, compassion, and joy to everyone.
Compassion	5,713	2.0%	76.8%	Strengthen your forcefield of peace and calmness by recognizing that others are suffering and cultivating the sincere wish that they be free from it.
Engaging Your Senses	5,658	2.0%	78.8%	Ground yourself in the present by bringing your attention to each of the senses, noticing and experiencing things as they are, without reacting to or judging them.
Counting Breaths	5,171	1.8%	80.6%	Slow the mind down when you need to focus on a certain task, when your thoughts are racing, or you're trying to fall asleep.
Change	5,087	1.8%	82.3%	Become more open and flexible with the recognition that everything is in a constant state of change, all of the time.
Nature Sounds	4,877	1.7%	84.0%	Take a time our by focusing on one thing-the sounds of nature-to de-stress and settle yourself down.
Lion Mind	4,710	1.6%	85.6%	Keep your mind on the bigger picture. Like a Lion, learn to notice distractions but not follow after them by focusing your attention on your breathing.
Commonality of Suffering	3,462	1.2%	86.8%	Find more patience, empathy and understanding as you learn to recognize how we're all in the boat.
Other	38,098	13.2%	100.0%	*Catch all for remaining MMAs*

Statistical Analysis

To determine the effect that a particular MMA has on the transition from one emotional state to another we used Generalized Linear Models (GLMs) as implemented in the R package lme4.[82] GLMs have many benefits which makes them suitable for our analysis. For example, GLMs can accommodate and quantify serial correlations among variables in longitudinal analyses. Additionally, both fixed and random effects can be considered for important covariates. Random effects are important to account for variance explained by unmeasured covariates such as individual-specific random effects of unknown origin (e.g., personal habits, unmeasured stressors or exposures, etc.). GLMs have also been widely used in statistical analysis of many psychiatric and psychological phenomena.[101-103] For each YMM quadrant we subset the data to users whose mood was associated with that quadrant when they initially engaged the app and then built four logit-link GLM models designed to predict which quadrant the user transitioned to after engaging in the MMA. To enable this, we created dummy variables to indicate if a user transitioned to a specific quadrant based. We used count data for the number of MMAs a user completed in a session as an independent variable, and also included age, sex, session index (i.e. 1 as the first use of the application, 2 as the second use, etc.) to capture repeated uses of the app, as additional predictors. Finally, we included other covariates, such as subscription status, user account completion, time between sessions, and country code in initial analyses, but were left out of subsequent analyses due to their insignificance and inability to contribute to predictions of transitions. Because of the differences in the number of user engagements (the range was from 10 uses to 1044 uses), we used a $\log_{10}$ transform on session index, which is consistent with what we did in our previous analysis of the SBT data. All non-MMA independent variables were standardized so that the resulting model beta values could be directly compared.

Modeling a Learning System

To evaluate the effectiveness of our analytical models in predicting transitions, we implemented the same methodology of fitting GLMs, but restricted our analyses to a 'training dataset' to three sequential observations per user starting with their first completed recorded session. After obtaining models for each

of the possible transitions based on data from these three initial sessions, we used the models to predict the probability of each transition in subsequent sessions and then selected the transition with the highest probability and matched it with the observed transitions. This allowed us to compare the actual emotional state transition with a predicted transition state and to see if the app could be improved by anticipating MMAs likely to result in positive mood transition in real time. We repeated this analysis using different numbers of and time-intervals in our training sets a number of times to further evaluate its performance.

Results

Dataset Summary

Prior to any filtering we had observations for nearly 5 million engagements with MMAs provided by the SBT app across 677,000 different users. There were 84,000 active users who completed 10 or more sessions who collectively completed 3.16 million engagements with MMAs provided by the app. After filtering for operating system, which users were active, language used in the app, and quality, 11,030 unique users were included in subsequent analyses. These users collectively completed 289,360 MMAs across 253,363 sessions (Average 1.14 MMAs per session). As shown in Table 3.2**,** most users were female (75.0%) and between the ages of 13 to 40. Compared to our previous publication we had fewer users and sessions, due to removal from legal and compliance, but the users for which we had data they completed more sessions and emotional check-ins on average. Additional population information can be found in Supplemental Figures 3.3-3.6.

Table 3.2 Study Population Statistics
Detailed study population statistics broken out by gender.

	Users Gender			
	Female	Male	Undeclared	All Users
count	8,274	1,814	942	11,030
	75.0%	16.4%	8.5%	100.0%
Average Age (±sd)	33.00	36.23	33.90	33.61
std	13.48	13.48	15.15	13.68
Avg. Completed Sessions (±sd)	38.59	51.97	40.74	40.97
std	55.33	78.15	65.63	60.76
Avg. Completed Pre & Post MMA Checkins (±sd)	21.29	30.91	22.43	22.97
std	37.86	60.16	42.61	42.87

Clustering Analysis of Emotions

The results of our clustering of emotions analysis were similar to those described in our previously published paper. The optimal number of clusters as defined by silhouette scores on PCoA and PAM analyses was eight. Of the 115 emotions, all but 3 emotions (Envious, Fiery, Self Critical) were assigned to the same clusters as in our previously published analyses. The eight clusters grouped emotions into categories with very common themes and were validated with prior SBT product internal and clinical groupings (see Figure 3.3 and Table 3.3). Based on these clusters a user's emotional state pre- and post-MMA, we could determine which category a user's emotions were most closely associated with using distances of selected emotional terms to cluster medoids.

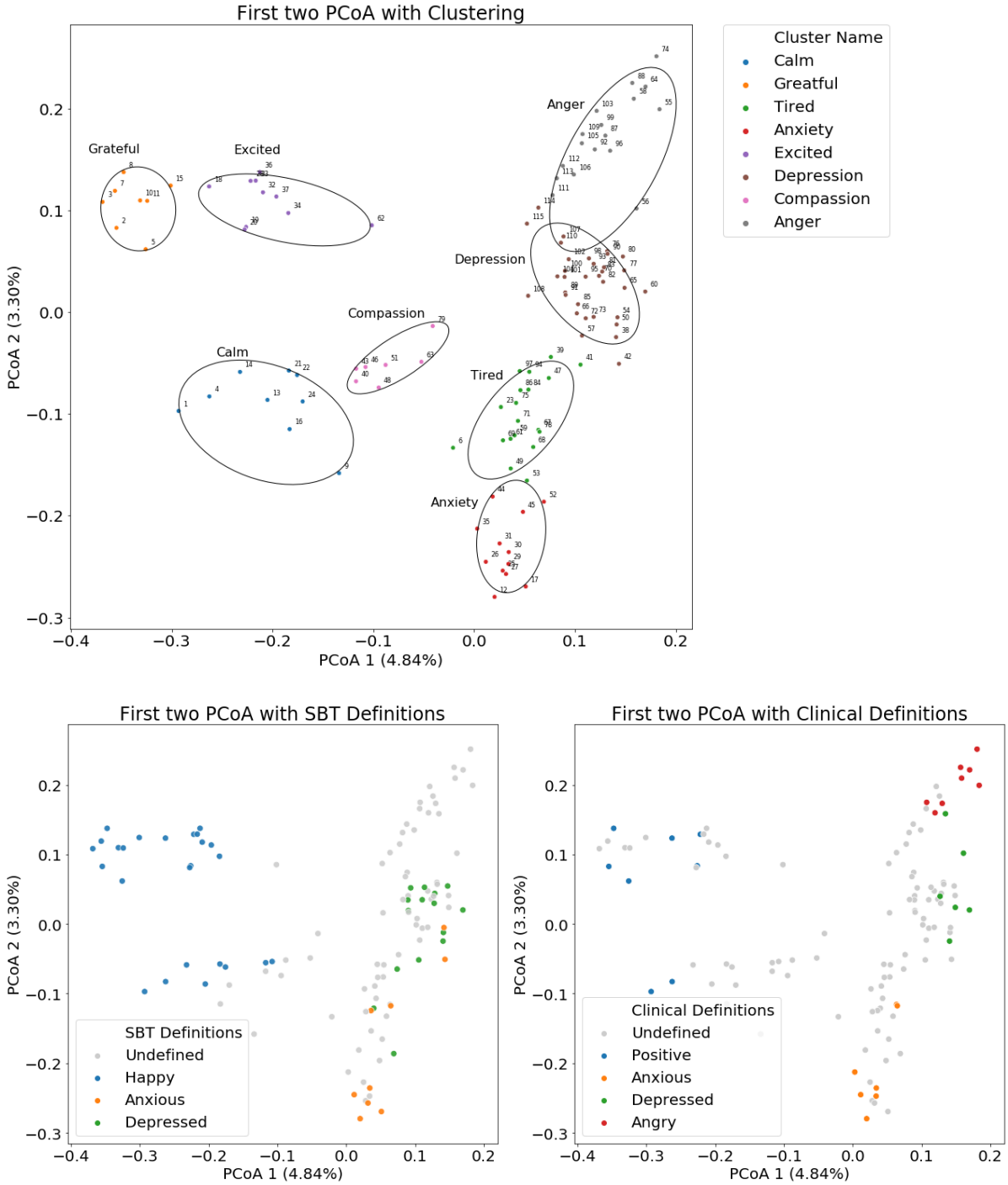


Figure 3.3 Emotional Clustering on Co-selected Terms
Emotional clusters created from co-selected terms within the same check-in, then defined using PAM and silhouette score. A) The optimal eight clusters are shown across the first two PCoAs. Clusters are given labels based on a single emotion that is thematic to most emotions within the cluster. B) In-house defined emotional labels show consistent grouping within the first two PCoAs C) Clinically defined emotional labels show consistent grouping within the first two PCoAs.

Table 3.3 Emotion Cluster and Quadrant Assignments
All 115 endorsable emotions grouped by cluster and associated YMM quadrant.

Assigned Cluster	YMM Mapping	Emotion
Calm	Green	Calm[1], Relaxed[4], Neutral[9], Grounded[13], Balanced[14], Open-Minded[16], Loving[21], Connected[22], Open-Hearted[24]
Grateful	Yellow	Content[2], Greatful[3], Hopeful[5], Appreciative[7], Happy[8], Thankful[10], Encouraged[11], Strong[15]
Tired	Blue	Tired[6], Achy[23], In Pain[39], Exhausted[41], Fatigue[47], Lazy[49], Indifferent[59], Confused[68], Hesitant[69], Bored[71], Insomnia[75], Apathetic[84]
	Red	Conflicted[53], Guarded[61], Afraid[67], Panicked[78], Torn[86], Clingy[94], Suspicious[97]
Anxiety	Blue	Distracted[31]
	Red	Anxious[12], Stressed[17], Scattered[25], Nervous[26], Uneasy[27], Restless[29], Worried[30], Concerned[35], Apprehensive[44], Mixed Up[45], Disconnected[52]
Excited	Yellow	Satisfied[18], Excited[19], Relieved[20], Enthusiastic[28], Proud[32], Fulfilled[33], Creative[34], Joyful[36], Glad[37], Fiery[62]
Depression	Blue	Sad[38], Overwhelmed[42], Lonely[50], Insecure[54], Sensitive[57], Depressed[60], Discouraged[65], Longing[66], Disappointed[70], Vulnerable[72], Lacking Confidence[73], Hurt[76], Powerless[77], Defeated[80], Heartbroken[81], Isolated[82], Unhappy[83], Not Confident[85], Grieving[89], Rejected[90], Guilty[91], Ashamed[93], Helpless[95], Despairing[98], Embarrassed[100], Despondent[101], Alienated[102], Remorseful[104], Homesick[108], Bullied[110], Competitive[114], Arrogant[115]
	Red	Insulted[107]
Compassion	Green	Caring[40], Kind[43], Compassionate[46], Non-Judgmental[48], Trusting[51], Empathetic[63], Equanimous[79]
Anger	Blue	Self Critical[56], Pessimistic[96], Defensive[99], Disgusted[109], Envious[111], Closed-Minded[113]
	Red	Frustrated[55], Annoyed[58], Irritated[64], Angry[74], Grumpy[87], Resentful[88], Impatient[92], Full Of Dislike[103], Judgmental[105], Jealous[106], Full Of Contempt[112]

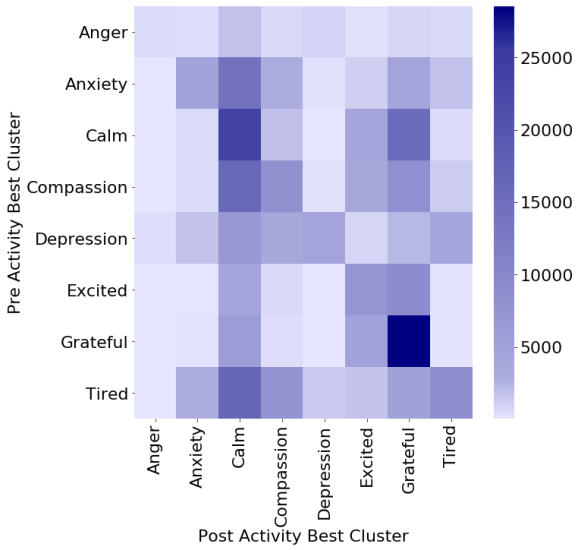
Alignment of clusters to YMM quadrants provided further support for the clustering, as most clusters were clearly aligned to a quadrant. As shown in Figure 3.4A, “Calm” and “Compassion” clusters had perfect alignment with the green quadrant, and “Grateful” and “Excited” clusters had perfect alignment with the yellow quadrant. The “Tired” cluster had emotions that crossed between both the blue (i.e. tired, lazy, fatigue) and red (i.e. afraid, panicked, suspicious) quadrants. The “Anger” cluster also had some crossover between red (i.e. angry, impatient, resentful) and blue (i.e. defensive, disgusted, pessimistic) quadrants. Conveniently, each quadrant was mapped to two clusters.

Figure 3.4B & 3.4C show relative frequency of pre-MMA emotional states and post-MMA emotional states. The most frequent ending cluster was calm, followed by grateful. User’s predominately started and ended in positive states (i.e., green/yellow clusters), with green (low energy, high pleasantness) being the predominant emotional state that users transitioned to. The most common negative states were tired and anxiety.

A) Cluster to Yale Mood Meter Assignment

Cluster	Blue	Green	Yellow	Red	Totals
1 Calm	0	9	0	0	9
2 Greatful	0	0	8	0	8
3 Tired	12	0	0	7	19
4 Anxiety	1	0	0	11	12
5 Excited	0	0	10	0	10
6 Depression	32	0	0	1	33
7 Compassion	0	7	0	0	7
8 Anger	6	0	0	11	17
Totals	51	16	18	30	**115**

B) Transitions by Cluster



C) Transitions YMM

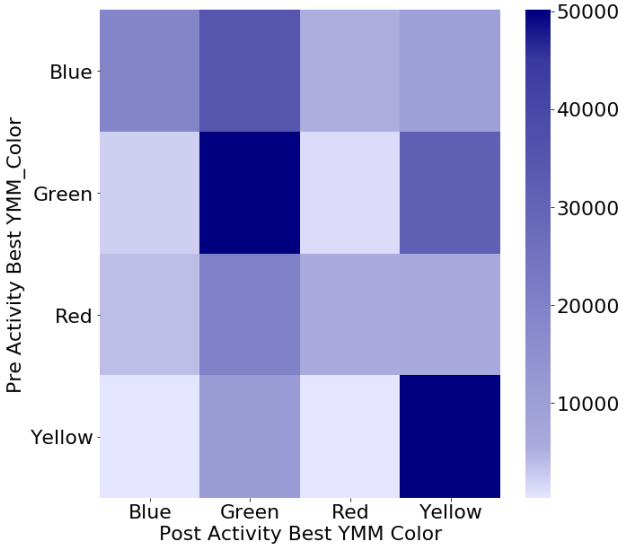


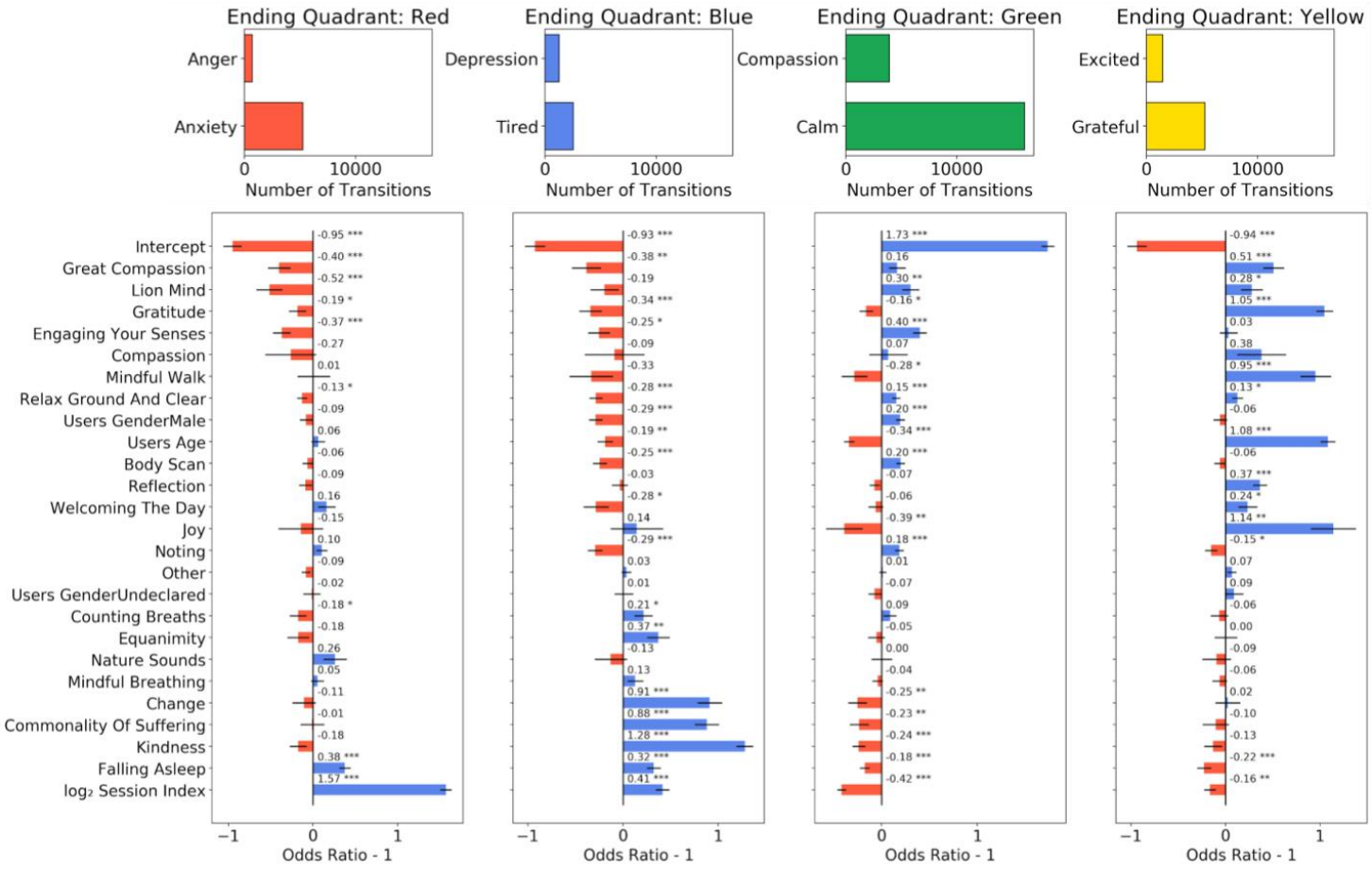
Figure 3.4 Cluster Alignment and Pre-to-post MMA Transitions
Clusters are assigned to YMM quadrants based on the majority of emotions within that cluster which correspond to the YMM quadrant (i.e. Tired cluster is Blue in YMM). A) Counts of all 115 endorsable emotions in the SBT app by each cluster and YMM quadrant they are associated with. B) Session counts for transitions from pre MMA emotional cluster to post MMA emotional state. Calm and grateful clusters are the most frequent ending states. C) Session counts for transitions from pre MMA emotional YMM quadrant to post MMA emotional quadrant.

Model Features

We did not find any specific MMA to be the most effective in transitioning users' emotions to different categories. Each of the sixteen GLMs we fit to the data centering on the effects of the different MMAs showed various degrees of association strength between pre and post MMA emotions. The MMAs *Great Compassion, Lion Mind,* and *Gratitude* provided by the app were all are associated with transitions from the red emotional state quadrant to a pleasant emotional state quadrant (green/yellow), whereas *Falling Asleep, Kindness,* and *Commonality of Suffering* were associated with staying in a negative emotional state. For users starting in the blue quadrant, the MMAs *Kindness*, *Great Compassion*, and *Counting Breaths* were associated with transitioning from green or yellow quadrants. Session index (a proxy for the number of engagements with the app over time) also was associated with users remaining or transitioning to low pleasantness quadrants. Both age and gender were associated with transitions of different sorts as well. For example, males were more likely than females to transition from negative states to positive states, and older users were more likely to transition to yellow states. There was not enough data to assess the degree to which MMAs could influence users who started in a yellow quadrant and ended in a red quadrant.

Figure 3.5 Transition Model Odds Ratios
Sixteen model fixed effect odd ratios and P values (<.001***, <.01**, and <.05*) for MMAs, gender, session index, and intercept split by starting YMM quadrant. Blue bars show increased propensity to make a transition, where-as red bars show decreased probability of making the transition. A) Transitions starting in Red YMM quadrant. B) Transitions starting in Blue YMM quadrant. C) Transitions starting in Green YMM quadrant. D) Transitions starting in Blue YMM quadrant. For example, looking at *A) Transitions from starting in Red YMM quadrant*, users who completed the MMA *Lion Mind* were 52% less likely to stay in Red, 30% more likely to enter Green, and 28% more likely to enter Yellow than those who did not complete *Lion Mind*.

A) Transitions from starting in Red YMM quadrant (Anger, Anxiety)



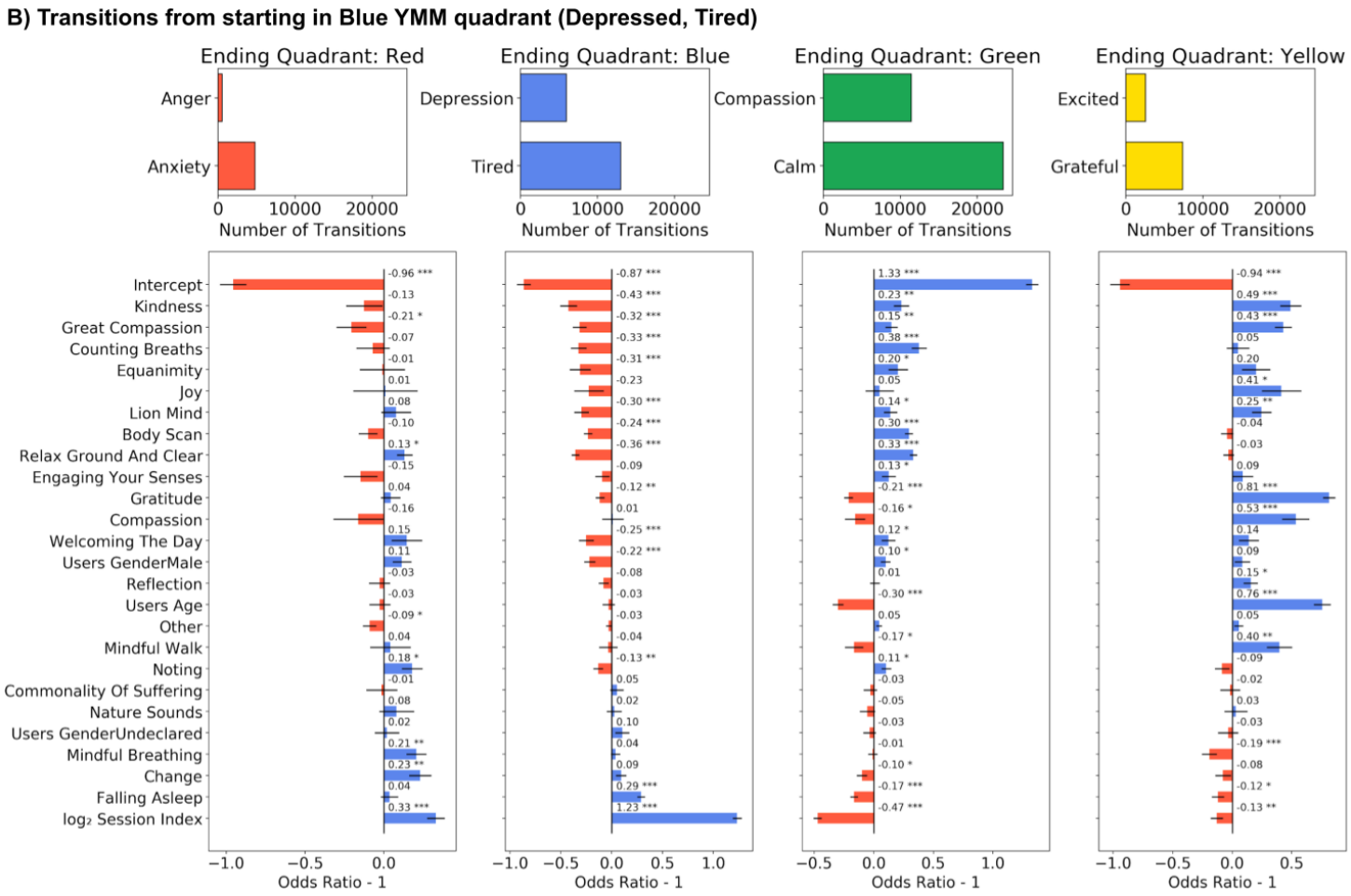


Figure 3.5 Transition Model Odds Ratios Continued

C) Transitions from starting in Green YMM quadrant (Compassion, Calm)

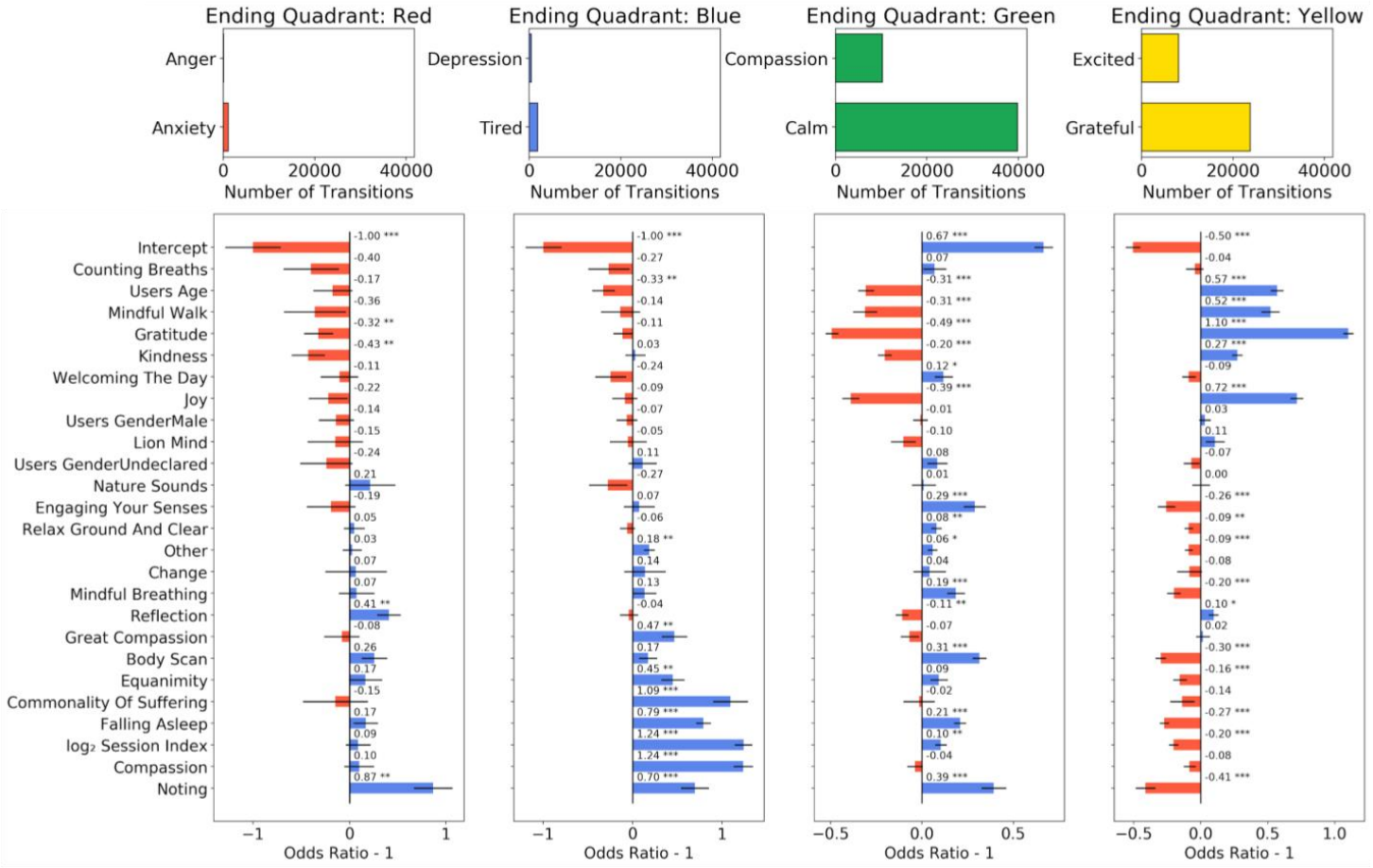


Figure 3.5 Transition Model Odds Ratios Continued

D) Transitions from starting in Green YMM quadrant (Excited, Grateful)

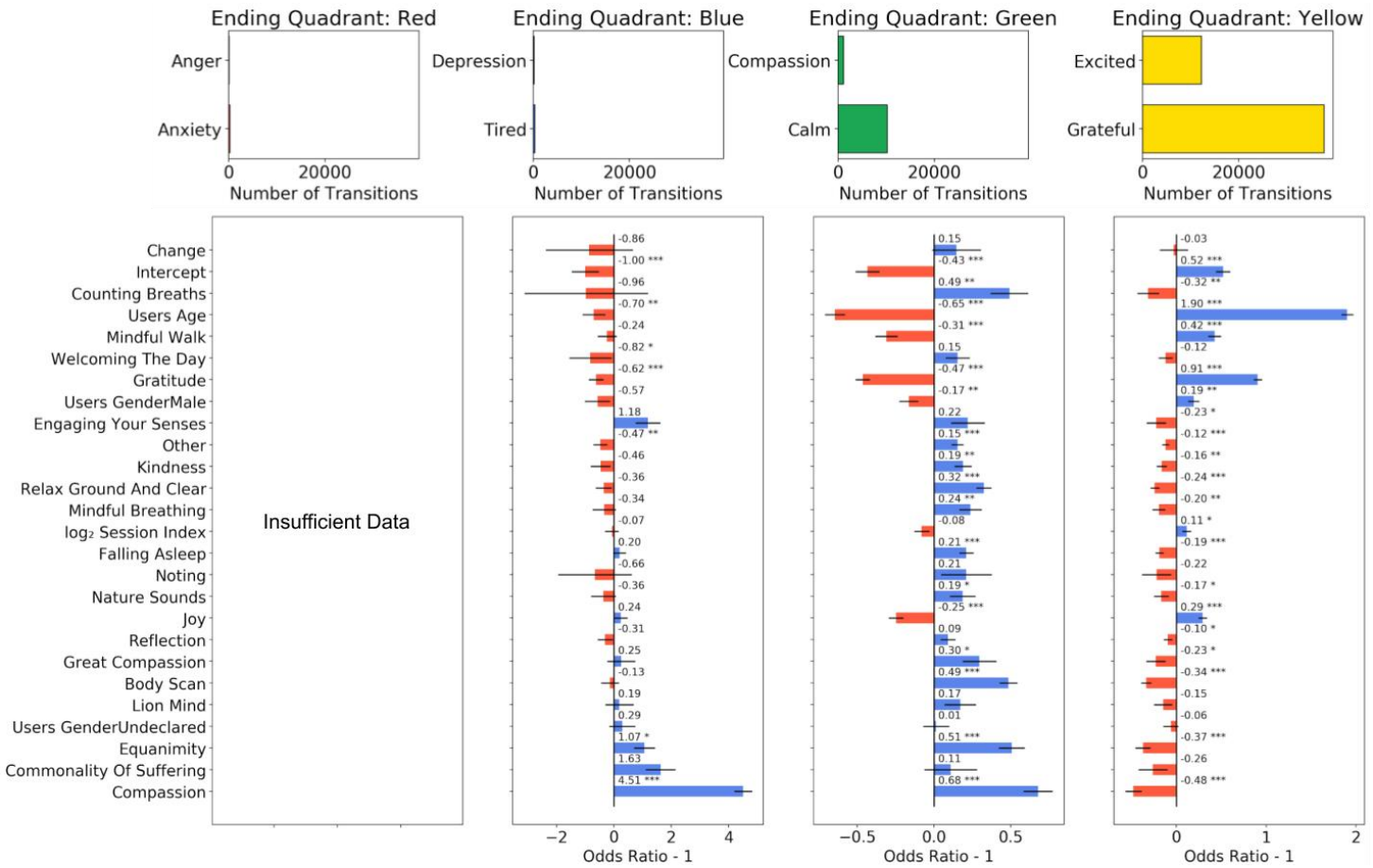


Figure 3.5 Transition Model Odds Ratios Continued

Rapid Learning System

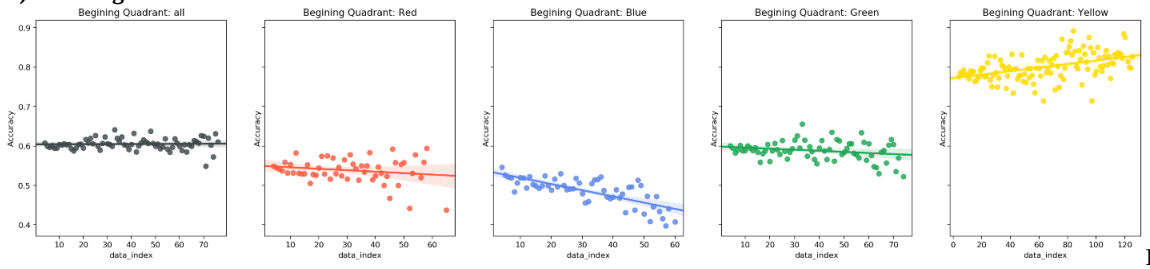
The most common quadrant for which users transitioned to was green, as expected for an application built to guide users to a calm and meditative state. The green emotional state quadrant accounted for ~45% of all emotional states that users transitioned to. Using data associated with the first 3 engagements with an SBT MMA, we built predictive models to determine which states users are most likely to transition to in subsequent engagements. We then compared the predicted transitions to the actual transitions. We found that the predictions were 61% accurate. Subsequent analyses involving the use of more data from users suggested that this accuracy level was maintained. We did find that the accuracy of the predictions differs as a function of the quadrant the emotions were assigned prior to pursuing the MMA, with blue being the least accurate, and yellow being the most accurate. We ultimately found that the more observations used in training the models, the more accurate the predictions become, as expected.

Table 3.4 Study Population Statistics

	Num Obs				
Ending State	3	4	5	10	15
Green	45.1%	44.9%	44.7%	44.0%	43.4%
Yellow	39.3%	39.5%	39.8%	40.7%	41.4%
Blue	10.4%	10.4%	10.4%	10.4%	10.4%
Red	5.2%	5.2%	5.1%	5.0%	4.8%
Overall Accuracy	60.7%	61.4%	61.7%	64.5%	65.4%

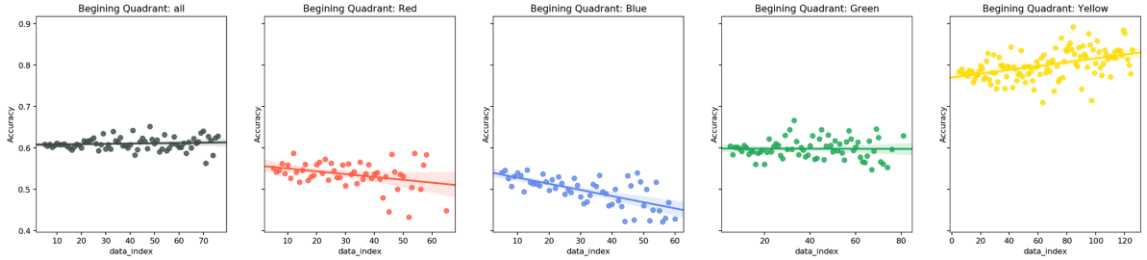
Figure 3.6 Ending YMM Quadrant Prediction Accuracy from Distance to Training Data
Accuracy of predictive models for ending YMM quadrant with varying number of observations used for training. Accuracy is also shown depending on the starting YMM quadrant which shows differing levels of accuracy and change in accuracy given the starting quadrant. A) 3 training observations. B) 4 training observations. C) 5 training observations. D) 5 training observations. E) 15 training observations

A) Training observations: 3

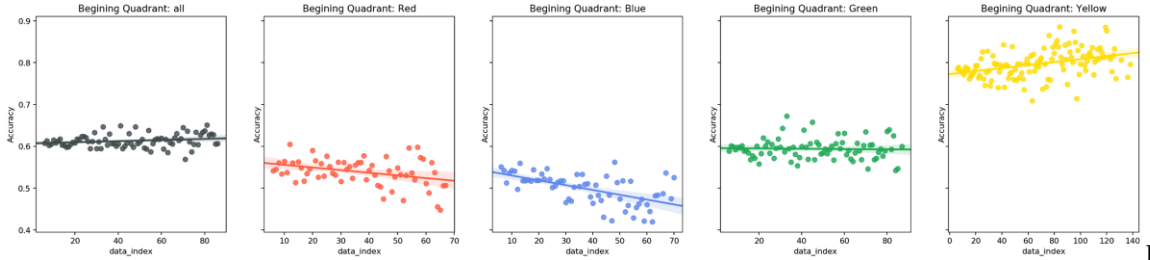


B)

Training Observations: 4

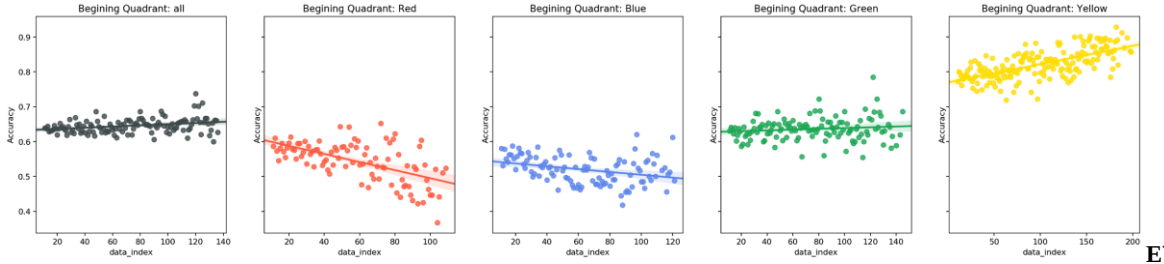


C) Training Observations: 5



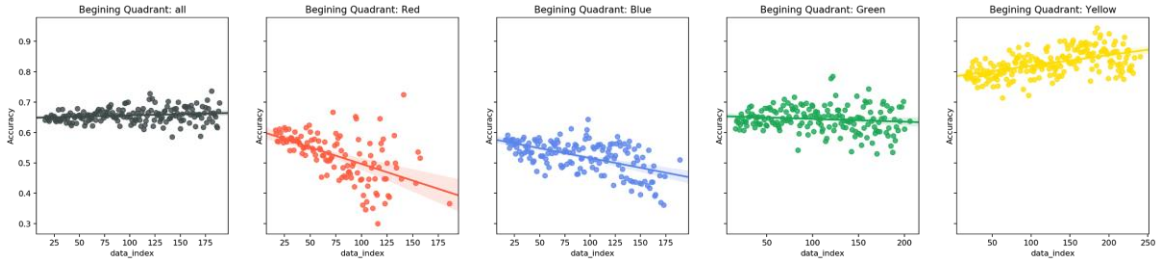
D)

Training Observations: 10



E)

Training Observations: 15



Discussion

Dataset

Due to the de-identification process of users, it is difficult to directly compare aspects of our current analyses with the results of our previously published studies. Despite this, we note that, in aggregate, the proportions of users in different emotional state categories was more or less the same (age, gender, emotion endorsement) between the studies. Our clustering performed analyses produced very similar results, which we interpret as consistency in the underlying data as well as in the final results. The key finding of the previous analysis, that users baseline emotional state improved with the number of engagements with the app helps put into perspective our findings from our most recent analyses.

Principal findings

Our analyses suggest that individual MMAs provided by the SBT app have varying degrees of influence on transitions between emotional states, based on a user's baseline emotional state. Furthermore, we find that there is no 'one-size fits all' solution to a user's poor mood in that not all MMAs affected individuals in the same way. Thus, we find evidence suggesting that depending on what a user's initial mood or mental state was at the time of an engagement with the SBT app prior to pursuing an MMA can influence whether or not a specific MMA will improve their mental state. Thus, with this in mind, for the app to be more effective, it would be important to 'nudge' users away from certain MMAs which might increase the probability that they will remain in a negative mental state. Gender differences also seem to play a role in how a user's mood will transition after engaging in an MMA, as males appear to have an easier transition away from unpleasant emotional states. A user's age also seemed to affect how they transition from poor mental states as well, as older users seem to heavily favor a more pleasant, energetic state.

Overall some clear themes emerged from analyses implicating each of our eight mental state clusters. The first PCoA resulting from our cluster analysis most strongly resembles the pleasantness axis

of the YMM, with the left-hand side being more pleasant clusters (Grateful, Calm, Compassion, Excited). The second PCoA somewhat resembles the energy axis of the YMM, with the clusters on the top having higher energy (Excited, Grateful, Angry); however, we found that the anxiety cluster (mapped to Red) was more ambiguous. This could suggest that an additional underlying emotional factor to anxiety exists that makes it different from other high energy, low pleasantness emotions. Additional projections (supplemental figure 3.7) with the 3rd and 4th PCoAs suggest that the angry and anxiety clusters share only weak similarity.

When we examined which MMAs drive emotional transitions, we see some common themes as well as a few surprising results. For example, the *Engaging Your Senses* MMA asks a user to tune into each of his or her senses in sequence, observing what they notice without evaluating or judging their experience. The ability to observe one's thoughts, rather than being fully caught up or entangled in them, has been referred to as 'metacognitive awareness,' and has been shown to be beneficial in dealing with anxiety and stress. This type of MMA may be ideally suited for producing specific transitions as our analyses suggest. Another MMA, *Great Compassion,* involves a 3 step process: 1) recognizing that others are just like you in that they want to experience happiness and avoid pain and suffering; 2) broadening one's attention to include people or pets that they love, people you don't know, and even people you have difficulty with, and then imagining you are breathing in pain and suffering, and breathing out positive energy; and 3) calling to mind people who are of service to others in the world who can inspire you to do the same. *Great Compassion* may have the effect of moving people out of an angry state because of the process of putting yourself in another's shoes, and then cultivating "big picture" thinking - i.e., looking at the world a little differently, from a broader perspective. The *Gratitude* MMA has a similar perspective and has the impact of reframing, and also helping to cultivate big picture thinking, helping to put things into a larger perspective. These three activities all have a focus on separating one's thoughts from the current emotions one is experiencing and thus are quite likely to lead to similar transitions.

The *Lion Mind* MMA, on the other hand, is a quieting activity, using the metaphor of lion mind vs. dog mind to help take one out of 'thought loops' (or ruminating thoughts) that feed anger. What is surprising is that the MMA *Kindness* seems ineffective, as it is thought of as the antidote to anger in the traditional Tibetan Buddhist perspective for which the MMA is founded on. It may be that this type of activity works better as a long-term remedy, but does not work as an immediate solution for, e.g., anger management. *Commonality of Suffering* and *Change* both are similar in that they help to put things into perspective by tapping into your empathy. With a broader perspective, it's supposed to be easier to feel more relaxed about your own situation or feelings, but users may not be able to reach this perspective given their emotional state. One additional complication may be that *Kindness* and *Change* use a somewhat different and more traditional way of communicating than other MMAs.

For transitions which start from the blue quadrant, we see less surprising results, but still notice some common themes. For MMAs which lead to a favorable transition, e.g., *Kindness*, *Great Compassion*, Counting *Breaths*, and *Equanimity*, there is a strong focus on interconnectedness and developing a bond to others. Using this information to anticipate needs for effecting changes in mental state, the app can better recommend and understand why some MMAs might be more effective given a user's emotional state.

Our previous analysis suggested that a user's baseline emotional state improves with continued use of the application. However, in this present study we noticed that the longer a user has engaged the app, the less likely they are to transition from a negative state, as defined by associations with literature-defined mood quadrants. While counterintuitive, these results do not contradict each other. We do see more users both starting and ending in green or yellow states (as shown in the learning system modeling studies). This observation might suggest there are some users who use the app but may not find it effective despite long term use, or there is a gradual "inoculation" of sorts whereby users don't find as much benefit from the app since they have benefitted maximally at some point in time. Additionally, there is a unique MMA, *Falling Asleep*, which is meant to help users relax and fall asleep. The 'tired' cluster can be both a negative, or

positive cluster depending on the circumstance, and users who select the *Falling Asleep* meditation could be moving to this more ambiguous cluster, driving some the results that we observed.

We find great potential in using the types of models we built to predict what outcome a user will have based on their MMA selections and initial mood or mental states. Even with a nominal number of observations we are able to accurately predict 60.7% of all transitions (15.6% increase over informed guessing, 35.7% increase over random guessing). These models get more accurate with more data, which suggests that a real-time learning system could be implemented to help guide users to MMAs which would have better chances of a successful transition. As more data is collected, further refinements to the predictive models could be made, as different covariates could be included in the analysis without the worry of overfitting. Other covariates could include, but are not limited to, environmental factors such as time since the last session, last MMA completed, current weather, political events, or other external events in a user's life.

Limitations of the Study

While our study was conducted with data from many users, each with several engagements with the app over varying periods of time, our study is entirely observational and does not include the sorts of controls in place in, e.g., randomized controlled clinical trials. Our analyses were also limited to data reflecting what the users of the app ultimately chose and disclosed within the app. Given that some MMAs have widely differing effects, this would suggest that there are certain MMAs which are likely better suited for inducing different transitions but should be studied in more tightly controlled settings. Note also that the emotional classifications themselves are experimental and there are many alternative concepts that may differ from our observations in terms of the YMM quadrants.[104]

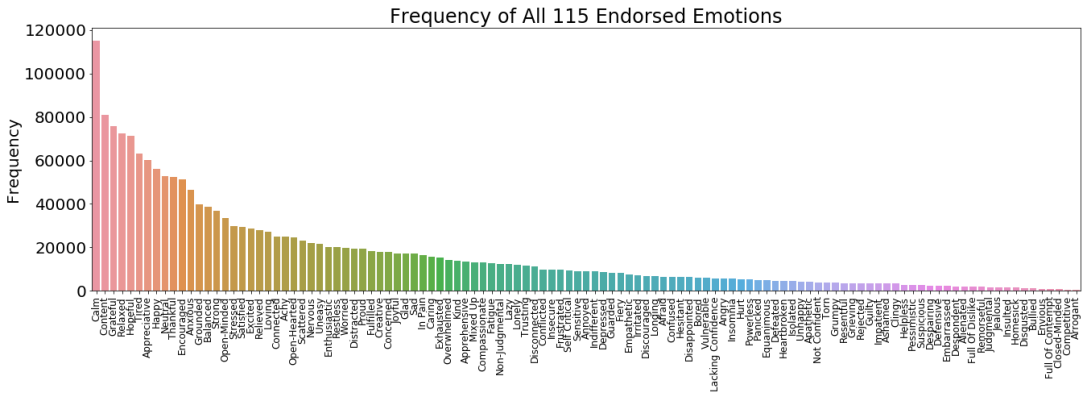
Our filtering criteria could have also created biases in our results. Since we examined individuals with 10 or more uses of the app, our attention was naturally confided to individuals that are engaged users

and have found some personal benefit for its continued use. In contrast, a user who stops after a few uses may not see the same benefits from MMAs, and hence not necessarily follow the observed transitions that long term users exhibit. Additionally, as noted in our previous Baseline Paper, long term use of the app influences basal emotional state. Finally, most users reported being in a YMM Green state (low energy – high pleasantness) when engaging the app initially and did not change their state post MMA, reducing the number of transitions we could study.

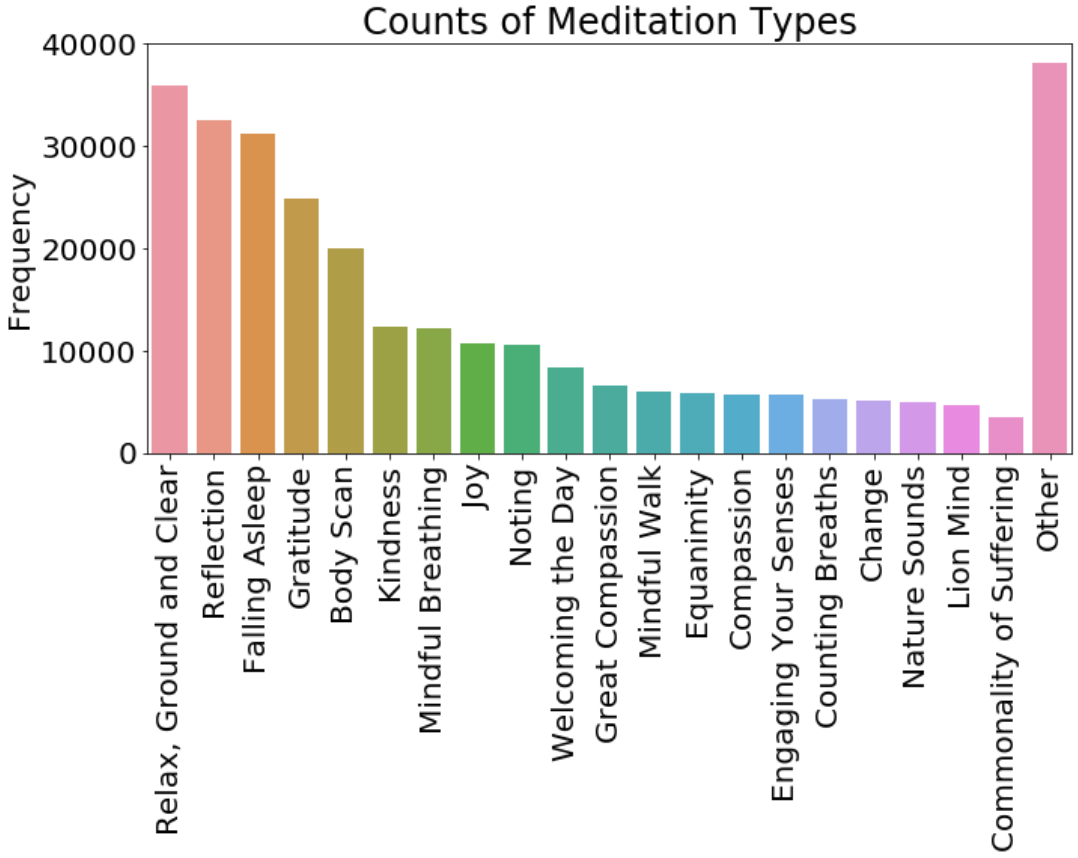
Future Directions

Since we confined our attention to specific users (e.g., iOS users) and MMAs (i.e., only those most widely used) we could expand our analyses to all users and MMAs, possibly by clustering the MMAs in some way. We focused our analysis on the transition from initial emotional state based on the chosen MMA but ignored other data that was collected (e.g., physical state of the person, sex, geolocation, etc.). We therefore could assess the degree to which these other factors impact our results. For example, we observed that males generally transition easier to improved mood more so than females (i.e., starting in a red state, males are 29% less likely to transition to blue and 20% more likely to transition to green), but we didn't test which MMAs work better for males (or females) individually. Knowing the effect of these factors, and how similar users react to MMAs, would help push our efforts towards the goal of truly personalizing the engagement and inducing a desired state of mind.

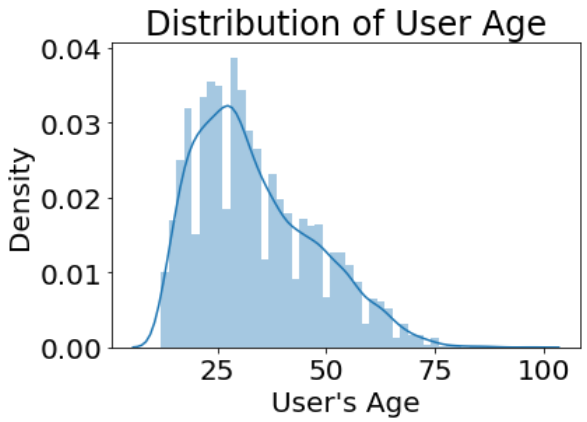
Supplemental Figures



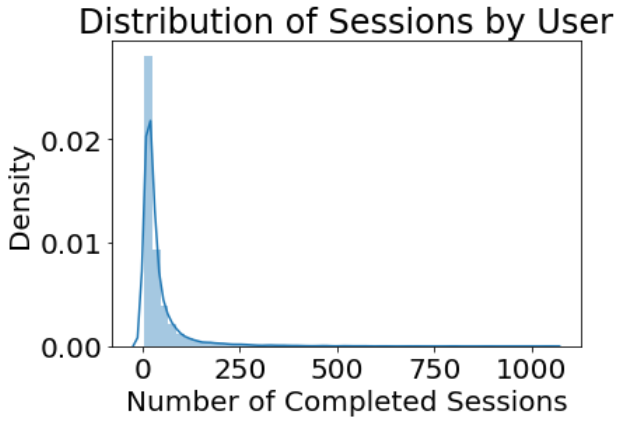
Supplemental Figure 3.1: Emotional Endorsement Frequency



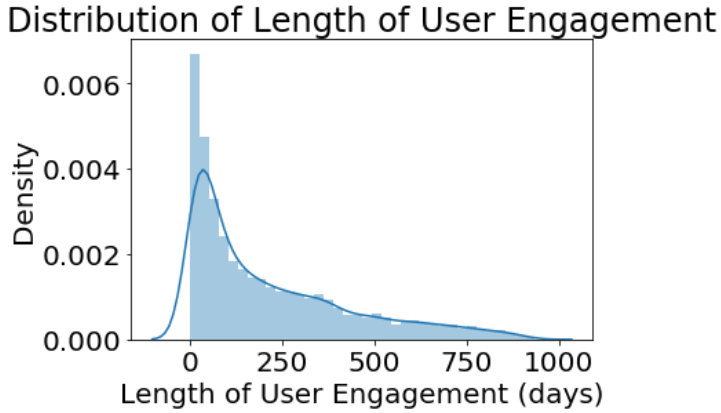
Supplemental Figure 3.2: Counts of MMAs completed



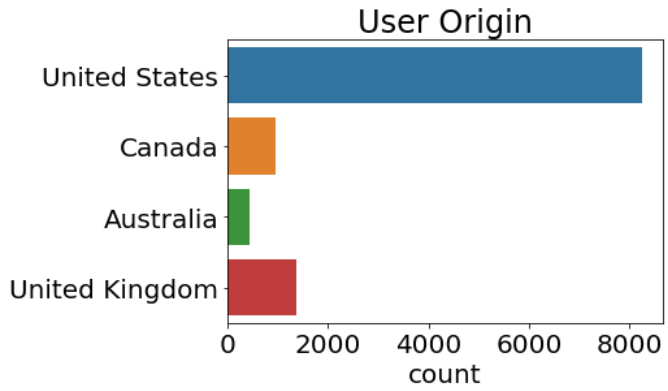
Supplemental Figure 3.3: Distribution of User Age



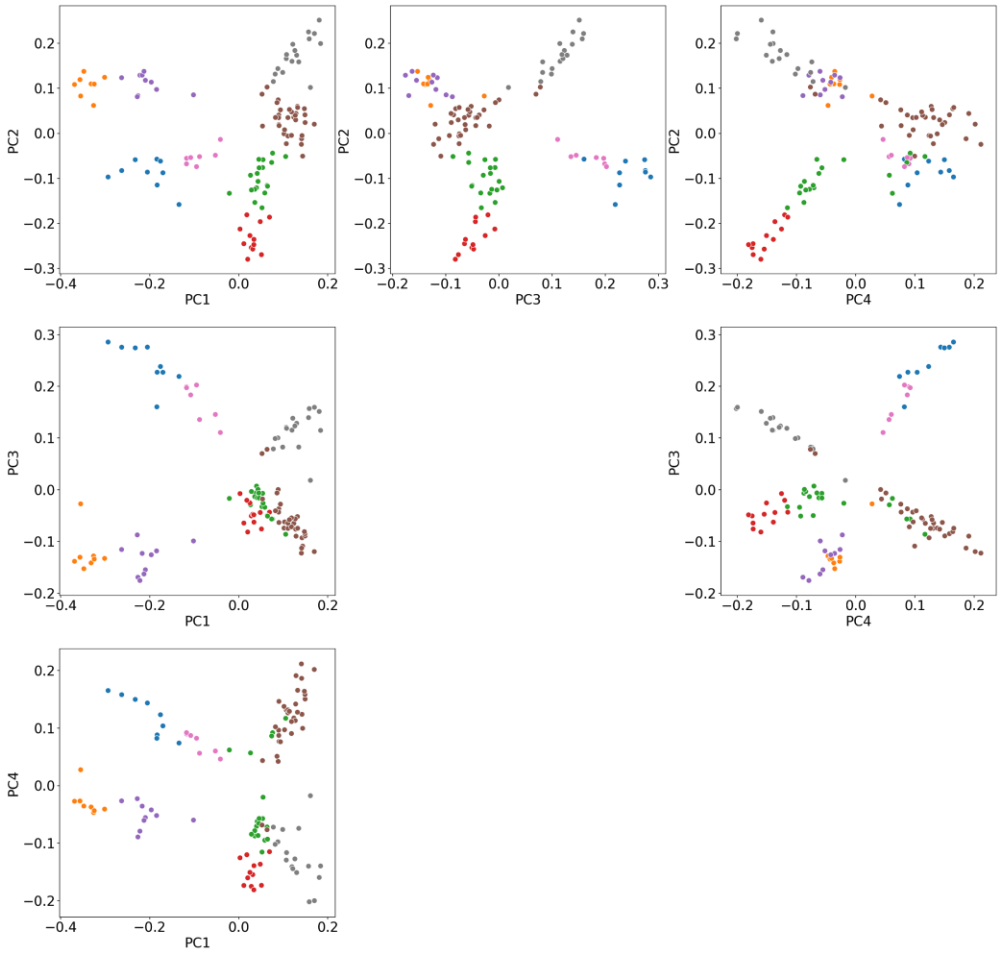
Supplemental Figure 4: Distribution of User Sessions



Supplemental Figure 5: Distribution of Length of Engagement



Supplemental Figure 6: User Origin



Supplemental Figure 7: Additional PCoA Projections

Chapter 3, in full, has been submitted for publication has been submitted for publication of the material as it may appear in Journal of Medical Internet Research, 2020, Argus J Athanas, Jamison McCorrison, Julie Campiston, Nick Bender, Jamie Price, Susan Smalley, Nicholas J Schork. The dissertation author was the primary investigator and author of this paper.

CHAPTER 4 DISCUSSION

Summary of Work

In each study we were able to use the extra dimensionality of longitudinal data to increase the power of detection, gain unique insights, or address otherwise impossible to answer questions. The works shown above highlight the importance of capturing and leveraging longitudinal data, and how it can be used to personalize medicine. This might come as risk factors in disease progression, assessing basal emotional state, or determining which therapies work best given your current disposition. Regardless of the domain, without capturing the trajectories of predictor variables we are missing a crucial component for developing personalized medicine.

Each dataset had its own unique and shared challenges to overcome. The Shiley patient data has several asynchronous measures on individuals for a multitude of tests. This was further confounded by non-independence of measures on each eye. The SBT data was complicated by multiple measures within the same time point, and both datasets involved missing, erroneous, and complex datatypes. These are just a few of the challenges that need to be addressed when using real world data, and if we are unable to do so it will be difficult to translate laboratory results to practice.

When considering the translation of these methods from trial to practice we must be cognizant of who is receiving and utilizing these results. A clinician or patient is likely not as technically savvy or willing to trust complete 'black box' algorithms. The implementation of standard and complex regression techniques provides an easier transition from theory to practice. An end user doesn't need to know what a beta or p value is to understand that concepts they stand for. For example, telling a patient that 'they are 20% more likely to experience an increased rate of visual field loss because of a genetic mutation,' is far more digestible than 'the machine learning algorithm says so.' With better patient education comes better adherence to the care plan and improved outcomes.

Future Directions

Much like a rapid learning system, these projects and methods could benefit from iterative improvement and there are several avenues for future investigation. While the papers themselves discuss some aspects of future directions they are not entirely inclusive. For the glaucoma data one of the biggest missed opportunities leveraging the longitudinal data is that for simplicity we collapsed the time series data into a single quantitative feature. Instead of fitting a simplistic linear model at each SNP we could again implore the use of LME models to use every timepoint. Because LMEs support hierarchical dependence structure, we could also use measurements from both eyes instead of just the worst one. This would allow us to create a regression model for each SNP which incorporates age, gene level admixture, intraocular pressure, medications, operation status, demographics, and comorbidities all while accounting for the inherited correlation from serial measures, right and left eye, and individual effects. The more features we can include within our models will allow us to explain more variance and have better power to detect true positives in our dataset. This approach is not without obstacles, first being with the data itself. Aligning the data is not trivial, and even once completed there are several ways of implementing the alignment. Determining which one is best is a mixture of domain knowledge and trial and error. While the dataset is deep, it is lacking in number of subjects, and importantly controls. With the number of confounding covariates (age, admixture, treatment, etc.) it is easy to overfit modeling for so few subjects, increasing the chance of false positives we end up seeing. Lastly a more complex model means more computation time. I found that running a single model with the data we had available for one SNP ran between 1 and 3 minutes. Extrapolating that to 6 million common SNPs across our subjects is 12 million minutes of compute time, or 22 years! Fortunately, this is not a linear problem, so distributive computing could reduce this to a couple of months in real time; however, this is still expensive (both in time and financially) and as noted before, there is likely some trial and error to create an optimal solution. This could also be addressed by writing some specialized software as the LME packages I was using in R are built to be flexible and may have some inefficiencies which a software engineer could overcome.

Throughout this project we have tried to incorporate more samples to increase our power or add viable controls. We tried to focus our inclusion criteria to WGS samples that had been sequenced and called using the same vendor and pipeline. Our first attempt was to include around 800 samples from the O'Connor study; however, much of the meta data for these samples was lost, and they were called using a different version of the reference genome. Ultimately the batch effects were too strong to include these samples. The second attempt occurred after giving a talk on our analysis and preliminary findings at ARVO 2018. We were approached by researchers from the UK who had access to roughly 3,000 additional samples that had variant calling done the same way as the ADAGES III dataset. We were never able to see if these samples could be used as we ran into bureaucratic issues with sharing personal data between countries. In the end, inclusion of more samples might help with providing a baseline and validation of previous known SNPs but is unlikely to have the same deep phenotyping on visual fields. Inclusion of more samples is not the only way to expand this dataset, as mentioned extensively, the power of the ADAGES III data is how subjects are deeply phenotyped over time. On top of visual field measures there is an abundance of imaging data which can be mined to access nerve fiber depth, vasculature, and abnormalities in the optic disk. Image data comes with its own problems, and handling and incorporating this data with genetics would be a very interesting problem that may provide some otherwise inaccessible insights.

Since the ADAGES III cohort is an ongoing trial, it lends itself as a perfect candidate for implementation and testing of a rapid learning system. The core principal is that predictions on an outcome are made and used to direct a care plan. Subsequent visits provide feedback if the treatment predictions are effective, and this new information is included in the model to redefine predictions. In a clinical setting treatment is entirely at the discretion of the clinician, but the model can still be used to still make predictions and evaluate effectiveness. Examining which features are driving prediction will provide information as to why some patients are progressing faster than others. In addition, various hierarchical clustering algorithms can be applied to subset patients into meaningful groups which can be individually studied for mechanisms

of action. Once a clear mechanism of action exists drug development becomes an easier and more straightforward task. While it is my belief that this should be standard practice with ongoing data collection, this same idea can be applied retroactively. To simulate continuous data collection, we can separate our data into test and training sets, not by random subject selection but by a fixed interval. For example, we would use the first three visits of every patient to train our rapid learning system and have it predict visual field decline. Then we would increment to the next time point and rebuild our models and predictions. This type of analysis would allow us to address several important questions: which features are most indicative of visual field decline, what is the relative effect of genetics vs. other clinical / demographic measures, the number of observations needed to make accurate predictions, and which models and assumptions provide the best predictive power.

The SBT analysis was performed on a dataset of users actively seeking treatment. To insure we had enough information on each user we subset our data such that only active users were included, which introduces some inherent biases. It is unlikely that people who the app didn't resonate with continued to use it. Like most real-world data, when you are doing active treatment it is difficult if not unethical to have true control patients where they are not receiving the standard of care. This means that we cannot directly attribute any results we see to the app or mindfulness and meditation activity. It could be useful to recruit healthy individuals and run a controlled experiment to determine the effects on baseline mood. This way we could attribute the application to changes more confidently. Additionally, we could expand this to try testing application use in targeted at risk groups such as, youths, socio-economic disadvantaged, or those with chronic medical conditions. This would have the benefits of allowing us to compare against the general population and determine if there are other useful meta-data that could be collected in order to make better MMA recommendations.

The SBT application and other digital therapeutics are perfect candidates for implementing an RLS. Because the care is delivered virtually and autonomously, the recommendation system is not on the

shoulders of clinician and instead dictated by an algorithm and acted upon by the individual receiving care. It is easy enough to version software and roll it out to subsets of your userbase. You can then in real time see which version of recommendation / prediction software is performing better for your users or patients. When you have enough confidence that the new system is working you can transition all users over to the same version of the app and then continue your iterative improvement. While there is still no true control the standard of care is being applied while new methods are being tested, and results can more directly be attributed changes in administration of the digital therapeutic.

These analyses are not limited to the datasets analyzed. There are many fields in which these longitudinal techniques, and rapid learning systems could be applied to real world data. The discordance between the two fields which I conducted my research is proof that they can be applied in many situations. Diseases like cancer, chronic kidney disease, and Alzheimer's where progressive decline is considered the most important endpoint are specific targets. Alzheimer's Disease (AD) is a progressive neurodegenerative disease which ultimately can result in death. AD comprises 60-70% of all dementia cases, and it is one of the largest public health concerns facing aging Americans.[3,105] While some medications show signs of slowing progression, AD, unfortunately, has no known cure. The ADNI is a longitudinal collaboration study which follows patients suspected of, or diagnosed with, AD. One of the specific aims of ADNI is to track progression with biomarkers. This makes it a great target to use similar methods as I conducted to explore the utility of genetics and clinical measures on predicting cognitive decline.

Current State of the Field

Risk calculators have been a common tool to help clinicians and patients both make informed decisions about what type of care they should get. While there are several studies which attempt to incorporate longitudinal factors into risk, longitudinal data is not often used in practice.[106-109] This is in part because quality longitudinal data is difficult to obtain due to interoperability issues, differing metrics, and

lack of recorded data. We have seen some specialties which have been able to successfully incorporate this type of data in real time for the benefit of the patient. For instance, with diabetes, where patients are using wearable insulin pumps this information can be relayed wirelessly and instantaneously and results in better patient outcomes.[110] Some diabetics have taken this one step further and made their own N-of-1 RLS by hacking their insulin monitoring and administration device.[111] Overall this is still a growing field and needs to be a focus of upcoming research to bring the use of longitudinal data into real world settings for calculating risk factors and predictions. For glaucoma specifically there has been some increased interest in doing more of these types of analyses. Since the time of my initial analysis (early 2018) there have been several publications which have also attempted to answer the same question as my analysis, although none of which have had the same deeply phenotyped data or the same approach.[112-114] A common approach has been to reduce SNPs into a polygenic risk score instead of testing individual SNPs or looking at gene burden. There is no way of knowing which approach will be the best until we have enough data, but my intuition says that is will be a mixture of domain specific knowledge and using robust statistical tests that can leverage time series data.

The concepts of rapid learning systems have been around since at least the 1970s when the field of biomedical informatics was just forming.[115] Today we still see many challenges which impede the implementation of such systems, even though they are still often thought of as the holy grail of personalized medicine.[116] We have seen many research hospitals make moves towards developing RLS especially in oncology.[117,118] City of Hope is one of the institutions which is trying to pave the way by integrating genomics with clinical outcome data, but they admit the process of doing this is still on the frontier of medicine. Some major hurdles yet to overcome include those of interoperability between providers, versioning of both software and references, accurate collection of outcomes, and increasingly policy concerns. It is my belief that interoperability between healthcare systems will always exist until strict data standardizations are in place. ICD 10 codes and the initial HITECH mandate are good starting points, but a stricter backbone for healthcare data normalization needs to exist before institutions can freely share data.

As we gain more information or have the ability to collect new information, the features we may want to include in a rapid learning system might change. While this is a tenant of any RLS it is also a challenge because it introduces the possibility of missingness within your data. In addition, some standards we treat as constants are bound to change as well. Within my research, the new standard for the human reference genome slowly switch from GRCh37 to GRCh38. At the start of my work with the Shiley data this caused innumerous problems as most tools were still set for using the older GRCh37 while the data was aligned and called using the newest version of the reference genome. Technology companies have been successful at making sure their applications are backwards compatible for at least a few versions, otherwise every time someone got a new phone, or upgraded their operating system none of their applications would work. Serious thought must be considered on how best to implement similar 'backwards' compatibility within genomics and medicine. Lastly, we must also consider the ramifications sharing information may have on an individual's privacy. For personalized medicine to be effective it must encompass everything that is important to the individual, including their right to anonymity.

REFERENCES

1 Ray, K. K., Kastelein, J. J., Boekholdt, S. M., Nicholls, S. J., Khaw, K. T., Ballantyne, C. M., Catapano, A. L., Reiner, Z. & Luscher, T. F. The ACC/AHA 2013 guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular disease risk in adults: the good the bad and the uncertain: a comparison with ESC/EAS guidelines for the management of dyslipidaemias 2011. *Eur Heart J* **35**, 960-968, doi:10.1093/eurheartj/ehu107 (2014).

2 Page, R. C., Martin, J., Krall, E. A., Mancl, L. & Garcia, R. Longitudinal validation of a risk calculator for periodontal disease. *J Clin Periodontol* **30**, 819-827, doi:10.1034/j.1600-051x.2003.00370.x (2003).

3 Barnes, D. E. & Yaffe, K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol* **10**, 819-828, doi:10.1016/S1474-4422(11)70072-2 (2011).

4 Mansberger, S. L. A risk calculator to determine the probability of glaucoma. *J Glaucoma* **13**, 345-347, doi:10.1097/00061198-200408000-00014 (2004).

5 Bailey, J. N., Loomis, S. J., Kang, J. H., Allingham, R. R., Gharahkhani, P., Khor, C. C., Burdon, K. P., Aschard, H., Chasman, D. I., Igo, R. P., Jr., Hysi, P. G., Glastonbury, C. A., Ashley-Koch, A., Brilliant, M., Brown, A. A., Budenz, D. L., Buil, A., Cheng, C. Y., Choi, H., Christen, W. G., Curhan, G., De Vivo, I., Fingert, J. H., Foster, P. J., Fuchs, C., Gaasterland, D., Gaasterland, T., Hewitt, A. W., Hu, F., Hunter, D. J., Khawaja, A. P., Lee, R. K., Li, Z., Lichter, P. R., Mackey, D. A., McGuffin, P., Mitchell, P., Moroi, S. E., Perera, S. A., Pepper, K. W., Qi, Q., Realini, T., Richards, J. E., Ridker, P. M., Rimm, E., Ritch, R., Ritchie, M., Schuman, J. S., Scott, W. K., Singh, K., Sit, A. J., Song, Y. E., Tamimi, R. M., Topouzis, F., Viswanathan, A. C., Verma, S. S., Vollrath, D., Wang, J. J., Weisschuh, N., Wissinger, B., Wollstein, G., Wong, T. Y., Yaspan, B. L., Zack, D. J., Zhang, K., Study, E. N., Consortium, A., Weinreb, R. N., Pericak-Vance, M. A., Small, K., Hammond, C. J., Aung, T., Liu, Y., Vithana, E. N., MacGregor, S., Craig, J. E., Kraft, P., Howell, G., Hauser, M. A., Pasquale, L. R., Haines, J. L. & Wiggs, J. L. Genome-wide association analysis identifies TXNRD2, ATXN2 and FOXC1 as susceptibility loci for primary open-angle glaucoma. *Nat Genet* **48**, 189-194, doi:10.1038/ng.3482 (2016).

6 Zhao, J., Feng, Q., Wu, P., Lupu, R. A., Wilke, R. A., Wells, Q. S., Denny, J. C. & Wei, W. Q. Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction. *Sci Rep* **9**, 717, doi:10.1038/s41598-018-36745-x (2019).

7 Barak-Corren, Y., Castro, V. M., Javitt, S., Hoffnagle, A. G., Dai, Y., Perlis, R. H., Nock, M. K., Smoller, J. W. & Reis, B. Y. Predicting Suicidal Behavior From Longitudinal Electronic Health Records. *Am J Psychiatry* **174**, 154-162, doi:10.1176/appi.ajp.2016.16010077 (2017).

8 Singh, A., Nadkarni, G., Gottesman, O., Ellis, S. B., Bottinger, E. P. & Guttag, J. V. Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration. *J Biomed Inform* **53**, 220-228, doi:10.1016/j.jbi.2014.11.005 (2015).

9 Garcia, T. P. & Marder, K. Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington's Disease as a Model. *Curr Neurol Neurosci Rep* **17**, 14, doi:10.1007/s11910-017-0723-4 (2017).

10 Fitzmaurice, G. M. & Ravichandran, C. A primer in longitudinal data analysis. *Circulation* **118**, 2005-2010, doi:10.1161/CIRCULATIONAHA.107.714618 (2008).

11 Rast, P. & Hofer, S. M. Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. *Psychol Methods* **19**, 133-154, doi:10.1037/a0034524 (2014).

12 Edwards, L. J. Modern statistical techniques for the analysis of longitudinal data in biomedical research. *Pediatr Pulmonol* **30**, 330-344, doi:10.1002/1099-0496(200010)30:4<330::aid-ppul10>3.0.co;2-d (2000).

13 Lu, N., Han, Y., Chen, T., Gunzler, D. D., Xia, Y., Lin, J. Y. & Tu, X. M. Power analysis for cross-sectional and longitudinal study designs. *Shanghai Arch Psychiatry* **25**, 259-262, doi:10.3969/j.issn.1002-0829.2013.04.009 (2013).

14 Rudrapatna, V. A., & Butte, A. J. . Opportunities and challenges in using real-world data for health care. *The Journal of Clinical Investigation* **130**, 565-574 (2020).

15 Franklin, J. M. & Schneeweiss, S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clin Pharmacol Ther* **102**, 924-933, doi:10.1002/cpt.857 (2017).

16 FDA. *FDA's Sentinel Initiative*, <http://www.fda.gov/safety/fdas-sentinel-initiative.> (October 18, 2019).

17 Carrigan, G., Whipple, S., Capra, W. B., Taylor, M. D., Brown, J. S., Lu, M., Arnieri, B., Copping, R. & Rothman, K. J. Using Electronic Health Records to Derive Control Arms for Early Phase Single-Arm Lung Cancer Trials: Proof-of-Concept in Randomized Controlled Trials. *Clin Pharmacol Ther* **107**, 369-377, doi:10.1002/cpt.1586 (2020).

18 Yue, L. Q. Leveraging Real-World Evidence Derived from Patient Registries for Premarket Medical Device Regulatory Decision-Making. *Statistics in Biopharmaceutical Research* **10**, 98-103, doi:10.1080/19466315.2017.1422436 (2018).

19 Lademann, M., Lademann, M., Boeck Jensen, A. & Brunak, S. Incorporating symptom data in longitudinal disease trajectories for more detailed patient stratification. *International Journal of Medical Informatics* **129**, 107-113, doi:https://doi.org/10.1016/j.ijmedinf.2019.06.003 (2019).

20 Zauderer, M. G. Practical Application of Real-World Evidence in Developing Cancer Therapies. *JCO Clinical Cancer Informatics*, 1-2, doi:10.1200/cci.19.00071 (2019).

21 Noah, B., Keller, M. S., Mosadeghi, S., Stein, L., Johl, S., Delshad, S., Tashjian, V. C., Lew, D., Kwan, J. T., Jusufagic, A. & Spiegel, B. M. R. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. *NPJ Digit Med* **1**, 20172, doi:10.1038/s41746-017-0002-4 (2018).

22 Stewart, M., Norden, A. D., Dreyer, N., Henk, H. J., Abernethy, A. P., Chrischilles, E., Kushi, L., Mansfield, A. S., Khozin, S., Sharon, E., Arunajadai, S., Carnahan, R., Christian, J. B., Miksad, R. A., Sakoda, L. C., Torres, A. Z., Valice, E. & Allen, J. An Exploratory Analysis of Real-World End Points for Assessing Outcomes Among Immunotherapy-Treated Patients With Advanced Non-Small-Cell Lung Cancer. *JCO Clin Cancer Inform* **3**, 1-15, doi:10.1200/CCI.18.00155 (2019).

23 Shah, A., Stewart, A. K., Kolacevski, A., Michels, D. & Miller, R. Building a Rapid Learning Health Care System for Oncology: Why CancerLinQ Collects Identifiable Health Information to Achieve Its Vision. *J Clin Oncol* **34**, 756-763, doi:10.1200/JCO.2015.65.0598 (2016).

24 Schilsky, R. L., Michels, D. L., Kearbey, A. H., Yu, P. P. & Hudis, C. A. Building a rapid learning health care system for oncology: the regulatory framework of CancerLinQ. *J Clin Oncol* **32**, 2373-2379, doi:10.1200/JCO.2014.56.2124 (2014).

25 Shrager, J. & Tenenbaum, J. M. Rapid learning for precision oncology. *Nat Rev Clin Oncol* **11**, 109-118, doi:10.1038/nrclinonc.2013.244 (2014).

26 Etheredge, L. M. A rapid-learning health system. *Health Aff (Millwood)* **26**, w107-118, doi:10.1377/hlthaff.26.2.w107 (2007).

27 Miotto, R., Wang, F., Wang, S., Jiang, X. & Dudley, J. T. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics* **19**, 1236-1246, doi:10.1093/bib/bbx044 (2017).

28 Tran, V. T., Riveros, C. & Ravaud, P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. *NPJ Digit Med* **2**, 53, doi:10.1038/s41746-019-0132-y (2019).

29 Zangwill, L. M., Ayyagari, R., Liebmann, J. M., Girkin, C. A., Feldman, R., Dubiner, H., Dirkes, K. A., Holmann, M., Williams-Steppe, E., Hammel, N., Saunders, L. J., Vega, S., Sandow, K., Roll, K., Slight, R., Auerbach, D., Samuels, B. C., Panarelli, J. F., Mitchell, J. P., Al-Aswad, L. A., Park, S. C., Tello, C., Cotliar, J., Bansal, R., Sidoti, P. A., Cioffi, G. A., Blumberg, D., Ritch, R., Bell, N. P., Blieden, L. S., Davis, G., Medeiros, F. A., Ng, M. C. Y., Das, S. K., Palmer, N. D., Divers, J., Langefeld, C. D., Freedman, B. I., Bowden, D. W., Christopher, M. A., Chen, Y. I., Guo, X., Taylor, K. D., Rotter, J. I., Weinreb, R. N., African, D. & Glaucoma Evaluation Study, I. I. I. G. S. G. The African Descent and Glaucoma Evaluation Study (ADAGES) III: Contribution of Genotype to Glaucoma Phenotype in African Americans: Study Design and Baseline Data. *Ophthalmology* **126**, 156-170, doi:10.1016/j.ophtha.2017.11.031 (2019).

30 Schork, A. J., Schork, M. A. & Schork, N. J. Genetic risks and clinical rewards. *Nat Genet* **50**, 1210-1211, doi:10.1038/s41588-018-0213-x (2018).

31 Springelkamp, H., Iglesias, A. I., Cuellar-Partida, G., Amin, N., Burdon, K. P., van Leeuwen, E. M., Gharahkhani, P., Mishra, A., van der Lee, S. J., Hewitt, A. W., Rivadeneira, F., Viswanathan, A. C., Wolfs, R. C., Martin, N. G., Ramdas, W. D., van Koolwijk, L. M., Pennell, C. E., Vingerling, J. R., Mountain, J. E., Uitterlinden, A. G., Hofman, A., Mitchell, P., Lemij, H. G., Wang, J. J., Klaver, C. C., Mackey, D. A., Craig, J. E., van Duijn, C. M. & MacGregor, S. ARHGEF12 influences the risk of glaucoma by increasing intraocular pressure. *Hum Mol Genet* **24**, 2689-2699, doi:10.1093/hmg/ddv027 (2015).

32 Chen, Y., Lin, Y., Vithana, E. N., Jia, L., Zuo, X., Wong, T. Y., Chen, L. J., Zhu, X., Tam, P. O., Gong, B., Qian, S., Li, Z., Liu, X., Mani, B., Luo, Q., Guzman, C., Leung, C. K., Li, X., Cao, W., Yang, Q., Tham, C. C., Cheng, Y., Zhang, X., Wang, N., Aung, T., Khor, C. C., Pang, C. P., Sun, X. & Yang, Z. Common variants near ABCA1 and in PMM2 are associated with primary open-angle glaucoma. *Nat Genet* **46**, 1115-1119, doi:10.1038/ng.3078 (2014).

33 Wiggs, J. L., Yaspan, B. L., Hauser, M. A., Kang, J. H., Allingham, R. R., Olson, L. M., Abdrabou, W., Fan, B. J., Wang, D. Y., Brodeur, W., Budenz, D. L., Caprioli, J., Crenshaw, A., Crooks, K., Delbono, E., Doheny, K. F., Friedman, D. S., Gaasterland, D., Gaasterland, T., Laurie, C., Lee, R. K., Lichter, P. R., Loomis, S., Liu, Y., Medeiros, F. A., McCarty, C., Mirel, D., Moroi, S. E., Musch, D. C., Realini, A., Rozsa, F. W., Schuman, J. S., Scott, K., Singh, K., Stein, J. D., Trager, E. H., Vanveldhuisen, P., Vollrath, D., Wollstein, G., Yoneyama, S., Zhang, K., Weinreb, R. N., Ernst, J., Kellis, M., Masuda, T., Zack, D., Richards, J. E., Pericak-Vance, M., Pasquale, L. R. & Haines, J. L. Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma. *PLoS Genet* **8**, e1002654, doi:10.1371/journal.pgen.1002654 (2012).

34 Medeiros, F. A., Zangwill, L. M., Bowd, C., Mansouri, K. & Weinreb, R. N. The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. *Invest Ophthalmol Vis Sci* **53**, 6939-6946, doi:10.1167/iovs.12-10345 (2012).

35 Bussel, II, Wollstein, G. & Schuman, J. S. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol* **98 Suppl 2**, ii15-19, doi:10.1136/bjophthalmol-2013-304326 (2014).

36 Leung, C. K. Diagnosing glaucoma progression with optical coherence tomography. *Curr Opin Ophthalmol* **25**, 104-111, doi:10.1097/ICU.0000000000000024 (2014).

37 Trikha, S., Saffari, E., Nongpiur, M., Baskaran, M., Ho, H., Li, Z., Tan, P. Y., Allen, J., Khor, C. C., Perera, S. A., Cheng, C. Y., Aung, T. & Vithana, E. A Genetic Variant in TGFBR3-CDC7 Is Associated with Visual Field Progression in Primary Open-Angle Glaucoma Patients from Singapore. *Ophthalmology* **122**, 2416-2422, doi:10.1016/j.ophtha.2015.08.016 (2015).

38 Nima Pouladi , K. B., Haiquan Li , Nancy G. Casanova , Tong Zhou , Wei Zhang , Ting Wang , Ikbel Achour , Rick Kittles , Yves Lussier , Joe G.N. Garcia. in *American Thoracic Society International Conference Abstracts* (2016).

39 Zhang, G., Saito, R. & Sharma, K. A metabolite-GWAS (mGWAS) approach to unveil chronic kidney disease progression. *Kidney Int* **91**, 1274-1276, doi:10.1016/j.kint.2017.03.022 (2017).

40 Selinski, S. Impact of urinary bladder cancer risk variants on prognosis and survival. *EXCLI J* **13**, 1254-1258 (2014).

41 Joseph, A., Kosmas, C. E., Patel, C., Doll, H. & Asherson, P. Health-Related Quality of Life and Work Productivity of Adults With ADHD: A U.K. Web-Based Cross-Sectional Survey. *J Atten Disord* **23**, 1610-1623, doi:10.1177/1087054718799367 (2019).

42 Aitken, L. M., Burmeister, E., McKinley, S., Alison, J., King, M., Leslie, G. & Elliott, D. Physical recovery in intensive care unit survivors: a cohort analysis. *Am J Crit Care* **24**, 33-39; quiz 40, doi:10.4037/ajcc2015870 (2015).

43 Goetzel, R. Z., Hawkins, K., Ozminkowski, R. J. & Wang, S. The health and productivity cost burden of the "top 10" physical and mental health conditions affecting six large U.S. employers in 1999. *J Occup Environ Med* **45**, 5-14, doi:10.1097/00043764-200301000-00007 (2003).

44 DALYs, G. B. D. & Collaborators, H. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and

territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1260-1344, doi:10.1016/S0140-6736(17)32130-X (2017).

45 Baer, R. A. Mindfulness Training as a Clinical Intervention: A Conceptual and Empirical Review. *Clinical Psychology-science and Practice* **6**, doi:DOI :10.1093/clipsy.bpg015 (2006).

46 Klingbeil, D. A., Renshaw, T. L., Willenbrink, J. B., Copek, R. A., Chan, K. T., Haddock, A., Yassine, J. & Clifton, J. Mindfulness-based interventions with youth: A comprehensive meta-analysis of group-design studies. *J Sch Psychol* **63**, 77-103, doi:10.1016/j.jsp.2017.03.006 (2017).

47 MacGregor, S., Ong, J. S., An, J., Han, X., Zhou, T., Siggs, O. M., Law, M. H., Souzeau, E., Sharma, S., Lynn, D. J., Beesley, J., Sheldrick, B., Mills, R. A., Landers, J., Ruddle, J. B., Graham, S. L., Healey, P. R., White, A. J. R., Casson, R. J., Best, S., Grigg, J. R., Goldberg, I., Powell, J. E., Whiteman, D. C., Radford-Smith, G. L., Martin, N. G., Montgomery, G. W., Burdon, K. P., Mackey, D. A., Gharahkhani, P., Craig, J. E. & Hewitt, A. W. Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma. *Nat Genet* **50**, 1067-1071, doi:10.1038/s41588-018-0176-y (2018).

48 Khor, Z. L. C. C. Current Development in Genome Wide Association Studies of Glaucoma. *Current Ophthalmology Reports volume* **6**, pages79–85 (2018).

49 Telenti, A., Pierce, L. C., Biggs, W. H., di Iulio, J., Wong, E. H., Fabani, M. M., Kirkness, E. F., Moustafa, A., Shah, N., Xie, C., Brewerton, S. C., Bulsara, N., Garner, C., Metzker, G., Sandoval, E., Perkins, B. A., Och, F. J., Turpaz, Y. & Venter, J. C. Deep sequencing of 10,000 human genomes. *Proc Natl Acad Sci U S A* **113**, 11901-11906, doi:10.1073/pnas.1613365113 (2016).

50 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. & Sham, P. C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

51 Kwon, Y. H., Kim, C. S., Zimmerman, M. B., Alward, W. L. & Hayreh, S. S. Rate of visual field loss and long-term visual outcome in primary open-angle glaucoma. *Am J Ophthalmol* **132**, 47-56, doi:10.1016/s0002-9394(01)00912-6 (2001).

52 Liebmann, K., De Moraes, C. G. & Liebmann, J. M. Measuring Rates of Visual Field Progression in Linear Versus Nonlinear Scales: Implications for Understanding the Relationship Between Baseline Damage and Target Rates of Glaucoma Progression. *J Glaucoma* **26**, 721-725, doi:10.1097/IJG.0000000000000710 (2017).

53 Hibar, D. P., Stein, J. L., Jahanshad, N., Kohannim, O., Hua, X., Toga, A. W., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Wright, M. J., Alzheimer's Disease Neuroimaging, I., Weiner, M. W. & Thompson, P. M. Genome-wide interaction analysis reveals replicated epistatic effects on brain structure. *Neurobiol Aging* **36 Suppl 1**, S151-158, doi:10.1016/j.neurobiolaging.2014.02.033 (2015).

54 M.J., C. *The R Book.* (John Wiley & Sons, 2012).

55 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. & Lee, J. J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

56 Wang, M., Huang, J., Liu, Y., Ma, L., Potash, J. B. & Han, S. COMBAT: A Combined Association Test for Genes Using Summary Statistics. *Genetics* **207**, 883-891, doi:10.1534/genetics.117.300257 (2017).

57 Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A. & Abecasis, G. R. A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

58 Choquet, H., Paylakhi, S., Kneeland, S. C., Thai, K. K., Hoffmann, T. J., Yin, J., Kvale, M. N., Banda, Y., Tolman, N. G., Williams, P. A., Schaefer, C., Melles, R. B., Risch, N., John, S. W. M., Nair, K. S. & Jorgenson, E. A multiethnic genome-wide association study of primary open-angle glaucoma identifies novel risk loci. *Nat Commun* **9**, 2278, doi:10.1038/s41467-018-04555-4 (2018).

59 Krumbiegel, M., Pasutto, F., Schlotzer-Schrehardt, U., Uebe, S., Zenkel, M., Mardin, C. Y., Weisschuh, N., Paoli, D., Gramer, E., Becker, C., Ekici, A. B., Weber, B. H., Nurnberg, P., Kruse, F. E. & Reis, A. Genome-wide association study with DNA pooling identifies variants at CNTNAP2 associated with pseudoexfoliation syndrome. *Eur J Hum Genet* **19**, 186-193, doi:10.1038/ejhg.2010.144 (2011).

60 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

61 Consortium, E. P. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9**, e1001046, doi:10.1371/journal.pbio.1001046 (2011).

62 Musiat, P. & Tarrier, N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. *Psychol Med* **44**, 3137-3150, doi:10.1017/S0033291714000245 (2014).

63 Insel, T. R. Assessing the economic costs of serious mental illness. *Am J Psychiatry* **165**, 663-665, doi:10.1176/appi.ajp.2008.08030366 (2008).

64 Smit, F., Cuijpers, P., Oostenbrink, J., Batelaan, N., de Graaf, R. & Beekman, A. Costs of nine common mental disorders: implications for curative and preventive psychiatry. *J Ment Health Policy Econ* **9**, 193-200 (2006).

65 Remes, O., Brayne, C., van der Linde, R. & Lafortune, L. A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain Behav* **6**, e00497, doi:10.1002/brb3.497 (2016).

66 Iacoviello, B. M., Steinerman, J. R., Klein, D. B., Silver, T. L., Berger, A. G., Luo, S. X. & Schork, N. J. Clickotine, A Personalized Smartphone App for Smoking Cessation: Initial Evaluation. *JMIR Mhealth Uhealth* **5**, e56, doi:10.2196/mhealth.7226 (2017).

67 Iacoviello, B. M., Murrough, J. W., Hoch, M. M., Huryk, K. M., Collins, K. A., Cutter, G. R., Iosifescu, D. V. & Charney, D. S. A randomized, controlled pilot trial of the Emotional Faces Memory Task: a digital therapeutic for depression. *NPJ Digit Med* **1**, doi:10.1038/s41746-018-0025-5 (2018).

68 Champion, L., Economides, M. & Chandler, C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: A pilot randomised controlled trial. *PLoS One* **13**, e0209482, doi:10.1371/journal.pone.0209482 (2018).

69 Lindsay, E. K., Chin, B., Greco, C. M., Young, S., Brown, K. W., Wright, A. G. C., Smyth, J. M., Burkett, D. & Creswell, J. D. How mindfulness training promotes positive emotions: Dismantling acceptance skills training in two randomized controlled trials. *J Pers Soc Psychol* **115**, 944-973, doi:10.1037/pspa0000134 (2018).

70 Economides, M., Martman, J., Bell, M. J. & Sanderson, B. Improvements in Stress, Affect, and Irritability Following Brief Use of a Mindfulness-based Smartphone App: A Randomized Controlled Trial. *Mindfulness (N Y)* **9**, 1584-1593, doi:10.1007/s12671-018-0905-4 (2018).

71 Bostock, S., Crosswell, A. D., Prather, A. A. & Steptoe, A. Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. *J Occup Health Psychol* **24**, 127-138, doi:10.1037/ocp0000118 (2019).

72 Coulon, S. M., Monroe, C. M. & West, D. S. A Systematic, Multi-domain Review of Mobile Smartphone Apps for Evidence-Based Stress Management. *Am J Prev Med* **51**, 95-105, doi:10.1016/j.amepre.2016.01.026 (2016).

73 Waltz, E. Pear approval signals FDA readiness for digital treatments. *Nat Biotechnol* **36**, 481-482, doi:10.1038/nbt0618-481 (2018).

74 Kvedar, J. C., Fogel, A. L., Elenko, E. & Zohar, D. Digital medicine's march on chronic disease. *Nat Biotechnol* **34**, 239-246, doi:10.1038/nbt.3495 (2016).

75 Harrison, V., Proudfoot, J., Wee, P. P., Parker, G., Pavlovic, D. H. & Manicavasagar, V. Mobile mental health: review of the emerging field and proof of concept study. *J Ment Health* **20**, 509-524, doi:10.3109/09638237.2011.608746 (2011).

76 Olff, M. Mobile mental health: a challenging research agenda. *Eur J Psychotraumatol* **6**, 27882, doi:10.3402/ejpt.v6.27882 (2015).

77 Blonde, L., Khunti, K., Harris, S. B., Meizinger, C. & Skolnik, N. S. Interpretation and Impact of Real-World Clinical Data for the Practicing Clinician. *Adv Ther* **35**, 1763-1774, doi:10.1007/s12325-018-0805-y (2018).

78 Berger, M. L., Sox, H., Willke, R. J., Brixner, D. L., Eichler, H. G., Goettsch, W., Madigan, D., Makady, A., Schneeweiss, S., Tarricone, R., Wang, S. V., Watkins, J. & Mullins, C. D. Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. *Value Health* **20**, 1003-1008, doi:10.1016/j.jval.2017.08.3019 (2017).

79 Monti, S., Grosso, V., Todoerti, M. & Caporali, R. Randomized controlled trials and real-world data: differences and similarities to untangle literature data. *Rheumatology (Oxford)* **57**, vii54-vii58, doi:10.1093/rheumatology/key109 (2018).

80 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290, doi:10.1093/bioinformatics/btg412 (2004).

81 Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927-930 (2003).

82 Douglas Bates, M. M., Ben Bolker, Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, doi:10.18637/jss.v067.i01 (2015).

83 A. P. Reynolds, G. R., B. de la Iglesia & V. J. Rayward-Smith Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. *Journal of Mathematical Modelling and Algorithms* **5**, 475–504, doi:https://doi.org/10.1007/s10852-005-9022-1 (2006).

84 Monto, M. A., McRee, N. & Deryck, F. S. Nonsuicidal Self-Injury Among a Representative Sample of US Adolescents, 2015. *Am J Public Health* **108**, 1042-1048, doi:10.2105/AJPH.2018.304470 (2018).

85 Xiao, H., Carney, D. M., Youn, S. J., Janis, R. A., Castonguay, L. G., Hayes, J. A. & Locke, B. D. Are we in crisis? National mental health and treatment trends in college counseling centers. *Psychol Serv* **14**, 407-415, doi:10.1037/ser0000130 (2017).

86 Ohrnberger, J., Fichera, E. & Sutton, M. The relationship between physical and mental health: A mediation analysis. *Soc Sci Med* **195**, 42-49, doi:10.1016/j.socscimed.2017.11.008 (2017).

87 Mark, T. L., Yee, T., Levit, K. R., Camacho-Cook, J., Cutler, E. & Carroll, C. D. Insurance Financing Increased For Mental Health Conditions But Not For Substance Use Disorders, 1986-2014. *Health Aff (Millwood)* **35**, 958-965, doi:10.1377/hlthaff.2016.0002 (2016).

88 Pohling, R., Buruck, G., Jungbauer, K. L. & Leiter, M. P. Work-related factors of presenteeism: The mediating role of mental and physical health. *J Occup Health Psychol* **21**, 220-234, doi:10.1037/a0039670 (2016).

89 Anxiety and Depression Association of America. Survey Finds that Americans Value Mental Health and Physical Health Equally. (2015).

90 Thomas, K. C., Ellis, A. R., Konrad, T. R., Holzer, C. E. & Morrissey, J. P. County-level estimates of mental health professional shortage in the United States. *Psychiatr Serv* **60**, 1323-1328, doi:10.1176/ps.2009.60.10.1323 (2009).

91 Butryn, T., Bryant, L., Marchionni, C. & Sholevar, F. The shortage of psychiatrists and other mental health providers: Causes, current state, and potential solutions. *International Journal of Academic Medicine* **3**, 5-9, doi:10.4103/ijam.ijam_49_17 (2017).

92 Capobianco, E. On Digital Therapeutics. *Frontiers in Digital Humanities* **2**, doi:10.3389/fdigh.2015.00006 (2015).

93 Sverdlov, O., van Dam, J., Hannesdottir, K. & Thornton-Wells, T. Digital Therapeutics: An Integral Component of Digital Innovation in Drug Development. *Clin Pharmacol Ther* **104**, 72-80, doi:10.1002/cpt.1036 (2018).

94 Terry, N. P. & Gunter, T. D. Regulating mobile mental health apps. *Behav Sci Law* **36**, 136-144, doi:10.1002/bsl.2339 (2018).

95 Torous, J. & Roberts, L. W. Needed Innovation in Digital Health and Smartphone Applications for Mental Health: Transparency and Trust. *JAMA Psychiatry* **74**, 437-438, doi:10.1001/jamapsychiatry.2017.0262 (2017).

96 Guerrero, A. P. S., Balon, R., Beresin, E. V., Louie, A. K., Coverdale, J. H., Brenner, A. & Roberts, L. W. Rural Mental Health Training: an Emerging Imperative to Address Health Disparities. *Acad Psychiatry* **43**, 1-5, doi:10.1007/s40596-018-1012-5 (2019).

97 Zoogman, S., Goldberg, S. B., Hoyt, W. T. & Miller, L. Mindfulness Interventions with Youth: A Meta-Analysis. *Mindfulness* **6**, 290-302, doi:10.1007/s12671-013-0260-4 (2015).

98 Azam, M. A., Latman, V. V. & Katz, J. Effects of a 12-Minute Smartphone-Based Mindful Breathing Task on Heart Rate Variability for Students With Clinically Relevant Chronic Pain, Depression, and Anxiety: Protocol for a Randomized Controlled Trial. *JMIR Res Protoc* **8**, e14119, doi:10.2196/14119 (2019).

99 Athanas, A. J., McCorrison, J. M., Smalley, S., Price, J., Grady, J., Campistron, J. & Schork, N. J. Association Between Improvement in Baseline Mood and Long-Term Use of a Mindfulness and Meditation App: Observational Study. *JMIR Ment Health* **6**, e12617, doi:10.2196/12617 (2019).

100 Nathanson, L., Rivers, S., Flynn, L. & Brackett, M. Creating Emotionally Intelligent Schools with RULER. *Emotion Review* **8**, doi:10.1177/1754073916650495 (2016).

101 Magezi, D. A. Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Front Psychol* **6**, 2, doi:10.3389/fpsyg.2015.00002 (2015).

102 Muth, C., Bales, K. L., Hinde, K., Maninger, N., Mendoza, S. P. & Ferrer, E. Alternative Models for Small Samples in Psychological Research: Applying Linear Mixed Effects Models and Generalized Estimating Equations to Repeated Measures Data. *Educ Psychol Meas* **76**, 64-87, doi:10.1177/0013164415580432 (2016).

103 Barr, D. J. Random effects structure for testing interactions in linear mixed-effects models. *Front Psychol* **4**, 328, doi:10.3389/fpsyg.2013.00328 (2013).

104 Lovheim, H. A new three-dimensional model for emotions and monoamine neurotransmitters. *Med Hypotheses* **78**, 341-348, doi:10.1016/j.mehy.2011.11.016 (2012).

105 Brookmeyer, R., Gray, S. & Kawas, C. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *Am J Public Health* **88**, 1337-1342, doi:10.2105/ajph.88.9.1337 (1998).

106 Balte, P., Cassano, P. A., Couper, D., Jacobs, D., Kalhan, R., Acien, A. N., Kaplan, R. C., Sanders, J., O'Connor, G. T., Schwartz, J. E., Yende, S., Dransfield, M. T., White, W. & Oelsner, E. C. in *C43. COPD AND POPULATION HEALTH* A4873-A4873.

107 Sticca, F., Ruggieri, S., Alsaker, F. & Perren, S. Longitudinal Risk Factors for Cyberbullying in Adolescence. *Journal of Community & Applied Social Psychology* **23**, 52-67, doi:10.1002/casp.2136 (2013).

108 Vaisto, J., Haapala, E. A., Viitasalo, A., Schnurr, T. M., Kilpelainen, T. O., Karjalainen, P., Westgate, K., Lakka, H. M., Laaksonen, D. E., Ekelund, U., Brage, S. & Lakka, T. A. Longitudinal associations of physical activity and sedentary time with cardiometabolic risk factors in children. *Scand J Med Sci Sports* **29**, 113-123, doi:10.1111/sms.13315 (2019).

109 Rueda-Ochoa, O. L., Smiderle-Gelain, M. A., Rizopoulos, D., Dhana, K., van den Berge, J.-K., Echeverria, L. E., Ikram, M. A., Deckers, J. W., Franco, O. H. & Kavousi, M. Risk factors for longitudinal changes in left ventricular diastolic function among women and men. *Heart* **105**, 1414-1422, doi:10.1136/heartjnl-2018-314487 (2019).

110 Rubin, M. P. a. R. R. Patient-Reported Outcomes for an Integrated Real-Time Continuous Glucose Monitoring/Insulin Pump System. *Diabetes Technology & Therapeutics* **11**, 57-62, doi:10.1089/dia.2008.0002 (2009).

111 Omer, T. Empowered citizen 'health hackers' who are not waiting. *BMC Med* **14**, doi:https://doi.org/10.1186/s12916-016-0670-y (2016).

112 Sakata, R., Yoshitomi, T., Iwase, A., Matsumoto, C., Higashide, T., Shirakashi, M., Aihara, M., Sugiyama, K., Araie, M. & Lower Normal Pressure Glaucoma Study Members in Japan Glaucoma, S. Factors Associated with Progression of Japanese Open-Angle Glaucoma with Lower Normal Intraocular Pressure. *Ophthalmology* **126**, 1107-1116, doi:10.1016/j.ophtha.2018.12.029 (2019).

113 Mabuchi, F., Mabuchi, N., Sakurada, Y., Yoneyama, S., Kashiwagi, K., Iijima, H., Yamagata, Z., Takamoto, M., Aihara, M., Iwata, T., Hashimoto, K., Sato, K., Shiga, Y., Nishiguchi, K. M., Nakazawa, T., Akiyama, M., Kawase, K., Ozaki, M., Araie, M. & Japan Glaucoma Society Omics, G. Genetic variants associated with the onset and progression of primary open-angle glaucoma. *Am J Ophthalmol*, doi:10.1016/j.ajo.2020.03.014 (2020).

114 Craig, J. E., Han, X., Qassim, A., Hassall, M., Cooke Bailey, J. N., Kinzy, T. G., Khawaja, A. P., An, J., Marshall, H., Gharahkhani, P., Igo, R. P., Jr., Graham, S. L., Healey, P. R., Ong, J. S., Zhou, T., Siggs, O., Law, M. H., Souzeau, E., Ridge, B., Hysi, P. G., Burdon, K. P., Mills, R. A., Landers, J., Ruddle, J. B., Agar, A., Galanopoulos, A., White, A. J. R., Willoughby, C. E., Andrew, N. H., Best, S., Vincent, A. L., Goldberg, I., Radford-Smith, G., Martin, N. G., Montgomery, G. W., Vitart, V., Hoehn, R., Wojciechowski, R., Jonas, J. B., Aung, T., Pasquale, L. R., Cree, A. J., Sivaprasad, S., Vallabh, N. A., consortium, N., Eye, U. K. B., Vision, C., Viswanathan, A. C., Pasutto, F., Haines, J. L., Klaver, C. C. W., van Duijn, C. M., Casson, R. J., Foster, P. J., Khaw, P. T., Hammond, C. J., Mackey, D. A., Mitchell, P., Lotery, A. J., Wiggs, J. L., Hewitt, A. W. & MacGregor, S. Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. *Nat Genet* **52**, 160-166, doi:10.1038/s41588-019-0556-y (2020).

115 Brown, J. H. The biomedical engineer and the health care system. *IEEE Trans Biomed Eng* **22**, 95-100, doi:10.1109/tbme.1975.324425 (1975).

116 Nass, S. J. & Balogh, E. National Cancer Policy Summit: addressing opportunities and challenges in cancer care and research. *Clin Pharmacol Ther* **89**, 770-771, doi:10.1038/clpt.2011.64 (2011).

117 Abernethy, A. P., Etheredge, L. M., Ganz, P. A., Wallace, P., German, R. R., Neti, C., Bach, P. B. & Murphy, S. B. Rapid-learning system for cancer care. *J Clin Oncol* **28**, 4268-4274, doi:10.1200/JCO.2010.28.5478 (2010).

118 Jones, R. D., Sabolch, A. N., Aakhus, E., Spence, R. A., Bradbury, A. R. & Jagsi, R. Patient Perspectives on the Ethical Implementation of a Rapid Learning System for Oncology Care. *J Oncol Pract* **13**, e163-e175, doi:10.1200/JOP.2016.016782 (2017).