Personalized prediction of response to smartphone-delivered meditation training: A machine learning approach

Christian A. Webb, PhD[1], Matthew J. Hirshberg, PhD[2],

Richard J. Davidson, PhD[2,3,4], Simon B. Goldberg, PhD[2,5]

[1]Harvard Medical School – McLean Hospital, Boston, MA, USA

[2]Center for Healthy Minds, University of Wisconsin – Madison, Madison, WI, USA

[3]Department of Psychology, University of Wisconsin – Madison, Madison, WI, USA

[4]Department of Psychiatry, University of Wisconsin – Madison, Madison, WI, USA

[5]Department of Counseling Psychology, University of Wisconsin – Madison, Madison, WI, USA

Corresponding authors:

Christian A. Webb, Ph.D.
Harvard Medical School & McLean Hospital
115 Mill Street, Belmont, MA 02478
Email: cwebb@mclean.harvard.edu

Simon B. Goldberg, Ph.D.
University of Wisconsin
335 Education Building, Madison, WI 53706
Email: sbgoldberg@wisc.edu

Abstract

Background: Meditation apps have surged in popularity in recent years, with an increasing number of individuals turning to these apps to cope with stress, including during the COVID-19 pandemic. In fact, meditation apps now represent the most commonly used mental health apps for depression and anxiety. However, little is known regarding who is well-suited to these apps.

Objective: The aim of this study was to develop and test a data-driven algorithm to predict which individuals are most likely to benefit from app-based meditation training.

Method: Using randomized controlled trial data comparing a 4-week meditation app (Healthy Minds Program; HMP) with an assessment-only control condition in school system employees (n = 662), we developed an algorithm predicting who is most likely to benefit from HMP. Baseline clinical and demographic characteristics were submitted to a machine learning model to develop a "Personalized Advantage Index" (PAI) reflecting an individual's expected reduction in distress (primary outcome) from HMP vs. control.

Results: A significant Group x PAI interaction emerged ($t(658) = 3.30$, $p = .001$), indicating that PAI scores moderated group differences in outcome. A regression model including repetitive negative thinking as the sole baseline predictor performed comparably well. Finally, we demonstrate the translation of a predictive model to personalized recommendations of expected benefit.

Conclusion: Overall, results reveal the potential of a data-driven algorithm to inform which individuals are most likely to benefit from a meditation app. Such an algorithm could be used to objectively communicate expected benefits to individuals, allowing them to make well-informed decisions about whether a meditation app is right for them.

Key words: precision medicine; prediction; machine learning; meditation; mobile technology; smartphone app.

Introduction

Precision medicine, which involves the use of individual variability to guide prevention and treatment, has gained momentum within the health sciences in the past several years [1]. This approach aspires to improve outcomes by matching patients with interventions most likely to yield success. In some medical specialties, precision medicine has led to impressive advances in personalized care. For example, research in oncology (e.g., lung and breast cancer) has effectively matched patients to targeted cancer treatments based on the unique genetic characteristics of their tumors, which has been shown to improve outcomes [2–4].

Psychiatry and clinical psychology have long hoped to better match patients with interventions. Numerous studies have examined patient-level factors (e.g., demographic, clinical and neurobiological characteristics) as predictors of treatment response [5,6]. However, with many potential predictors and inconsistencies across studies in the presence, direction, and strength of associations with outcome, empirically supported guidelines to inform optimal treatment matching remain elusive.

Machine learning has emerged as a promising analytical approach well-suited for handling and integrating large numbers of predictor variables, including correlated predictors, that may individually only modestly predict outcomes of interest but collectively can predict significant variance in patient outcomes [7,8]. Specific machine learning approaches such as decision-tree based algorithms (e.g., random forest) also effectively model non-linear and higher order interactions that may underlie predictive relationships [9]. In contrast to traditional statistical approaches which emphasize evaluating a specific hypothesis (i.e., null-hypothesis significance testing), machine learning models typically emphasize optimizing predictive performance, and evaluating the generalizability of models to new individuals (e.g., via cross-validation, holdout samples, or external validation) [10]. Machine learning approaches are increasingly being applied with some success within psychiatry and clinical psychology, with a growing number of studies demonstrating the ability to predict response to various psychiatric treatments [10–12].

In pursuit of precision mental health, researchers have leveraged machine learning approaches in an effort to optimize treatment recommendations [13–15]. For example, DeRubeis et al. [16] developed the *personalized advantage index* (PAI) as an algorithm for guiding treatment recommendations based on pretreatment patient characteristics. These models attempt to predict the benefit a specific patient would derive from Treatment A vs. Treatment B. The PAI has been successfully used to predict response to cognitive behavioral therapy (CBT) vs. an antidepressant medication [16], CBT vs. interpersonal therapy [17], CBT vs. psychodynamic therapy [18], and antidepressant medication vs. placebo [19].

Prior research using the PAI and related approaches [12] provide promising initial evidence that data-driven algorithms may improve patient outcomes by matching individuals to the most therapeutically beneficial treatment, as opposed to the current, suboptimal trial-and-error approach to treatment selection, which results in protracted psychiatric illness until an effective treatment is found. However, the fact remains that a substantial proportion of individuals suffering from a psychiatric disorder go untreated [20,21]. Digital health technologies, such as internet-based CBT [22] and smartphone-delivered mental health apps [23], have the potential to substantially increase access to evidence-based treatments [24]. Yet, the availability of thousands of mental health apps leave potential consumers faced with a dizzying number of choices with essentially no way of knowing which specific app may best suit their needs [25]. Data-driven treatment recommendation algorithms, such as the PAI, offer promising tools for informing optimal patient-treatment fit. Such approaches may also be valuable for addressing persistent limitations of mobile health (mHealth) approaches, including notoriously high and rapid disengagement [26,27]. Moreover, the scalability of mHealth makes collection of adequately powered sample sizes for robust modeling a possibility [28].

A recent analysis of available mental health apps revealed that meditation and mindfulness training (along with journaling and mood tracking) are the most common features offered across apps [29]. In fact, the two most widely used meditation apps (Headspace and Calm, with 5 million and 9 million monthly active users, respectively) account for 96% of daily active users in a recent evaluation of the top 27 apps for depression and anxiety [30]. Despite the soaring popularity of meditation apps, a critical question remains unanswered: *For whom* is app-based meditation training well-suited?

Current Study

The current study involved secondary analysis of a large-scale randomized controlled trial (RCT) comparing a meditation-based smartphone app, the Healthy Minds Program (HMP), with an assessment-only control condition [31]. The RCT was conducted in a sample of school district employees (n = 662) in the state of Wisconsin during the COVID-19 pandemic. Relative to pre-pandemic levels, rates of emotional distress and depressive symptoms increased substantially during the COVID pandemic [32]. Available evidence suggests that the emotional well-being of teachers also suffered during the pandemic [33,34], as they cope with COVID-related stressors, uncertainty and risks with the return to in-person instruction. Using data from the above RCT, the overarching goal of this study was to develop and evaluate a data driven (PAI) approach to inform personalized meditation app recommendations for school employees. Using readily gathered self-reported baseline demographic and clinical characteristics, we developed and tested a machine learning algorithm to identify which individuals are most likely to benefit from the HMP app.

Method

Participants and Procedure

Wisconsin school district employees were recruited via email and other electronic media between mid-June 2020 to late-August 2020 (for a full description of study procedures, see [31]). Eligible participants were adults (≥18 years of age) currently employed by a Wisconsin school who owned a smartphone capable of downloading the HMP, were fluent in English, had limited exposure to meditation or the HMP app, and depressive symptoms below the severe range (T-score < 70 on Patient-Reported Outcomes Information System [PROMIS] Depression [35]). T-scores are population normed at 50 with a standard deviation of 10. Upon completing pre-test measures, 666 participants were randomly assigned to use the four-week HMP or an assessment-only control condition (4 participants were removed for failing multiple attention checks; see Supplemental Figure S2 for CONSORT Flow Diagram). Participants completed weekly questionnaires during the intervention period (i.e., weeks 1, 2, 3) along with a post-treatment (week 4) and follow-up assessment (3-month after the end of the intervention period). These measures were administered via the web-based Research Electronic Data Capture (REDCap) survey system.

Study procedures were approved by the University of Wisconsin – Madison Institutional Review Board. The trial was preregistered at clinicaltrials.gov (NCT04426318) and through the Open Science Framework (https://osf.io/eqgt7). However, the current prediction analyses were not planned *a priori* and were not included in the preregistered data analytic plan. All code (implemented in the R statistical software) to reproduce the analyses in the manuscript has been posted on OSF (https://osf.io/94a6s/).

The HMP includes contemplative practices designed to build skills supportive of four pillars of well-being: awareness, connection, insight, and purpose [36,37]. Participants were encouraged to engage with content from each of the four modules for approximately one week (i.e., 4 weeks total). Content included didactic instruction as well as guided meditation practices. For the guided practices, participants could select the length of practice from 5 to 30 minutes. The HMP app was used for a mean of 10.9 days (SD = 9) over the 4-week trial. For additional trial and sample details, see [31].

Measures

Demographic characteristics. Participants reported their age, gender identity, race/ethnicity, marital status, and income at baseline.

Primary outcome. The prespecified primary outcome in the parent RCT was psychological distress which was a composite of the computer-adaptive versions of the PROMIS Anxiety and PROMIS Depression measures [35] and the 10-item Perceived Stress Scale (PSS) [38]. All three are widely used measures with established reliability and validity [39,40]. See Supplement for details. Consistent with the prespecified data analytic plan, multilevel models estimated changes in distress over the 4-week intervention period. Random slopes (representing individual change in distress over the intervention period) were calculated for each participant and served as the primary outcome in our machine learning prediction models.

Predictors. Several additional self-report questionnaires assessed secondary outcomes and candidate mediators theoretically linked to pillars of well-being trained within HMP. The 15-item Perseverative Thinking Questionnaire (PTQ) [41] assessed worry and rumination (α = .95).

The five-item World Health Organization (WHO)-5 [42] assessed global well-being (α = .85). The eight-item Act with Awareness subscale of the Five Facet Mindfulness Questionnaire (α = .91, FFMQ [43]) assessed mindful attention in daily life. The five-item National Institutes of Health (NIH) Toolbox Loneliness Questionnaire (NIHTL [44]) assessed perceived social disconnection (α = .90). The 12-item Self-Compassion Scale Short Form (SCSSF [45]) assessed feelings of kindness towards oneself (α = .86). The 10-item Drexel Defusion Scale (DDS [46]) assessed ability to experientially distance from internal experiences (α = .84). The 10-item Meaning in Life Questionnaire (MLQ [47]) assessed presence and search for meaning (αs = .91 and .93, respectively).

Analytic Strategy

Predictor variables included pre-intervention distress (composite measure), anxiety (PROMIS), depression (PROMIS), stress (PSS), repetitive negative thinking (PTQ), the mindfulness facet of acting with awareness (FFMQ), loneliness (NIHTL), defusion (DDS), presence (MLQ), search for meaning (MLQ), self-compassion (SCSSF), well-being (WHO-5), age, gender, race, marital status, and income.

Missing value imputation

Missing data were imputed using a random-forest based imputation (MissForest package in R [48]). To avoid contamination between predictor and outcome scores, which may optimistically bias predictive performance, the outcome variable (slope of change in distress) was excluded from the imputation procedure. Rates of missing data were very low, with no variable missing more than 6 values (6/662 = 0.9%). See Supplement for additional details.

Generating predicted outcomes

To predict outcomes, two prognostic models (using elastic net regularized regression; ENR; glmnet package in R) were developed, one for participants who received HMP and one for those who received the assessment-only control condition. To minimize overfitting which can occur with traditional *k*-fold cross-validation (CV), a nested CV procedure was used for each of these prognostic models (i.e., incorporating an outer and inner CV loop [49–52]). See Supplement for details on the nested CV procedure.

The above procedure generated predicted HMP outcomes for HMP participants, and predicted control condition outcomes for control participants. To generate predicted outcomes for the counterfactual condition (i.e., the treatment condition one did not receive), an ENR model was developed in one group (i.e., full HMP or control sample) and used to predict outcomes for participants in the other group.

Evaluation of Recommendations

As a final product of the above prediction models, every participant had two predicted outcome scores: one for HMP and one for the control condition. Consistent with prior similar studies [18,19,53], we computed a PAI score by subtracting these two predicted outcomes (i.e., predicted slope of change in distress for HMP *minus* control) for each individual. Thus, a negative PAI score indicates that a given participant is predicted to experience greater reductions (i.e., more negative slope) in distress in HMP relative to the assessment-only control condition (and vice versa for positive PAI scores). The PAI can be interpreted as a continuous indicator reflecting the expected magnitude of the advantage of one treatment condition relative to the other (e.g., a large negative PAI value indicates that the model predicts a relatively large between-group difference in outcome favoring HMP). We tested whether PAI scores moderated treatment group differences in outcome (i.e., slope of change in distress) via a Group (i.e., intervention condition) x PAI interaction. The latter test allowed us to answer the following question: Are more negative PAI scores (reflecting relatively greater predicted benefit from HMP relative to the control condition) in fact associated with larger *observed* differences in outcome favoring HMP?

Comparison Model

We compared the above multivariable machine learning (ENR) model with a simple linear regression with baseline repetitive negative thinking (PTQ) scores as the sole predictor (i.e., repeating the above steps to generate a PAI score for every participant), implemented via 10-fold CV (repeated 100 times to generate stable estimates). Repetitive negative thinking was selected as a predictor in this comparison model based on prior research indicating that it predicts response to mindfulness apps [51,54]. See Supplemental Materials for additional analyses with baseline distress as the sole predictor. Finally, we used the parameter estimates from the final models to demonstrate the translation of predicted outcomes to *personalized* recommendations for app-based mindfulness training.

All analyses were conducted in R (version 4.0.2) [55]. Sample size was originally determined for the purpose of the parent trial to detect between-group differences in the primary outcome (change in distress; see https://osf.io/eqgt7). To estimate whether the current sample size was adequately powered for the analyses proposed in the present study, a Monte Carlo simulation approach (InteractionPoweR package in R) was used. Informed by effect sizes from a prior mindfulness app RCT [54] which tested similar Group x PAI interactions, simulations revealed that a sample size of at least 153 was needed for Group x PAI interaction tests (with alpha = 0.05 and power = 80%; see Supplemental Figure S1, including Figure note, for additional power analysis details).

Results

Sample Demographics

The majority (79.0%) of participants reported depression and/or anxiety symptoms at baseline that were above the clinical cutoff on the PROMIS Depression and PROMIS Anxiety measures (T score > 55), and over half the sample (51.8%) reported moderate or greater anxiety or depressive symptoms at baseline (T > 60).

Groups did not differ at baseline on demographic or clinical variables (**Table 1**). Of those assigned to HMP, 95.6% downloaded the app and 78.8% used the app for one or more days. The mean number of days of use was 10.88 (SD = 9.08). The mean number of minutes of practice was 127.93 (SD = 130.63).

Outcome Prediction

Higher baseline levels of distress, depression and stress predicted better outcomes (i.e., greater reductions in distress) in HMP (see **Table 2**). The zero-order correlations between outcome and these three predictors were $r = -.30$ (for distress), $r = -.24$ (depression) and $r = -.30$ (stress). Predicted HMP outcomes were significantly correlated with observed outcomes for the HMP group ($r = .27$, $P < .001$; RMSE = 0.10), but not control condition outcomes ($r = .07$, $P = .21$; RMSE = 0.12). Conversely, predicted control condition outcomes were significantly correlated with observed outcomes for the control group ($r = .19$, $P < .001$; RMSE = 0.10), but not HMP outcomes ($r = .10$, $P = .06$; RMSE = 0.12). Higher baseline scores on the following variables predicted better outcome in the control condition: distress, anxiety, depression, stress, loneliness, defusion and presence. In addition, lower levels of repetitive negative thought, higher self-compassion and being married were each associated with better control condition outcome (**Table 2)**.

Meditation App Recommendations

The mean PAI score was -0.07 (SD = 0.03, range = -0.17 to 0.03) indicating that the model predicted greater average symptom improvement for the HMP meditation app relative to the assessment-only control condition. The model recommended HMP (PAI < 0) to all participants except 5 (657/662 = 99.2%).

Evaluation of Recommendations

A significant Group x PAI interaction emerged in predicting outcome ($t(658) = 3.30$, $P = .001$; adjusted $r^2 = .10$), indicating that PAI scores moderated group differences in outcome. As displayed in **Figure 1**, as PAI scores decrease (i.e., reflecting relatively stronger HMP recommendations) group differences in observed outcome increase, favoring HMP.

Comparison Model

For the linear regression comparison model applied to the HMP group, higher levels of repetitive negative thinking were significantly associated with a *greater* reduction in distress from the mindfulness app ($B = -0.02$, $t(342) = -3.37$, $P < .001$). The correlation between predicted HMP outcomes and observed outcomes was $r = .16$ ($P = .003$; RMSE = 0.10) for participants who received HMP and $r = -.14$ ($P = .015$; RMSE = 0.12) for the control group. In contrast to the pattern of findings for the HMP group, the linear regression model applied to the control sample revealed that higher levels of repetitive negative thinking were significantly associated with *poorer* outcomes to the control condition ($B = 0.01$, $t(316) = 2.44$, $P = .015$). The correlation between predicted control condition outcomes and observed outcomes was $r = .11$ ($P = .049$; RMSE = 0.11) for the control group and $r = -.18$ ($P < .001$; RMSE = 0.12) for the HMP group.

A significant Group x PAI interaction emerged in predicting change in distress ($t(658) =$

3.81, $P < .001$; adjusted $r^2 = .11$), indicating that PAI scores moderated group differences in outcome (**Figure 2)**. Specifically, as PAI scores decreased (reflecting increasing repetitive negative thinking scores) group differences favoring the HMP condition also increased. Given the association between repetitive negative thinking and depressive symptoms [56,57], we also conducted additional sensitivity analyses controlling for baseline levels of depressive symptoms (as well as considering number of days the app was used), which yielded the same pattern of findings (see Supplement). In sum, these results indicate that a simple linear regression including repetitive negative thinking as the sole predictor yields equivalent performance relative to a more complex multivariable ENR model (i.e., adjusted $r^2 = .11$ vs. $r^2 =.10$, respectively, for the Group x PAI interaction).

Translating a Predictive Model to Personalized Meditation App Recommendations

To demonstrate the translation of a predictive model to personalized recommendations, we used the parameter estimates from the above regression models to estimate predicted change in distress in HMP vs. the assessment-only condition for a new individual on the basis of their pre-intervention repetitive negative thinking score. Given that the simpler regression model performed similarly to the more complex multivariable ENR models, we used the former model for this demonstration.

First, as displayed in **Figure 3**, we plotted the relationship between PAI scores and outcome for HMP (blue line) and the assessment-only control condition (red line). The dashed vertical grey line represents the point at which the two regression lines intersect. An individual with a PAI score to the left of this line is predicted to have a better outcome in HMP relative to assessment only (and vice versa for individuals with PAI scores to right of this line). The area to the left of this line is colored yellow reflecting a “cautious recommendation” for app-based meditation training. Second, we computed a 95% confidence interval via bootstrap resampling (Boot package in R) [58]. Specifically, we drew 1,000 samples with replacement, and recomputed the two regression lines and their intersection point in each of these samples. The dashed vertical red line represents the left margin of the 95% confidence interval for this intersection point. In other words, if an individual's PAI score falls to the left of this line our confidence in the predicted benefit of HMP relative to the assessment-only condition increases. Third, we also implemented the Johnson-Neyman technique [59](Interactions package in R) to probe the Group x PAI interaction and to estimate the value of the moderator (PAI) at which group differences in outcome become statistically significant. This occurred at PAI < -0.02 (solid vertical grey line in **Figure 3,** immediately adjacent to the dashed red line). If a participant's PAI score falls to the left of both the 95% confidence interval (dashed red line) and the latter Johnson-Neyman threshold (solid grey line) the plot area is colored green to reflect a more confident recommendation to use HMP.

To illustrate with a concrete example, an individual with a repetitive negative thinking (PTQ) score one SD above the mean (i.e., 41) would have a PAI score of -0.10 (within the "green zone" of **Figure 3**), and a predicted slope of change in distress of -0.049 (i.e., expected reduction in distress) in HMP vs. 0.047 (i.e., expected increase in distress) in the assessment-only condition over 4 weeks. Assuming this individual had a pre-intervention level of distress at the 50th percentile, they would be predicted to be at the 41st percentile (relative to pre-intervention distress scores) following the 4-week mindfulness app course vs. 58th percentile if they only completed symptom assessments (i.e., control condition). In summary, based on a brief assessment of perseverative negative thinking, our algorithm can provide individual users with useful information regarding their expected benefit *prior* to them deciding to enroll in a multiweek course of app-based meditation training.

Discussion

An increasing number of individuals are turning to meditation apps to alleviate emotional distress. Indeed, meditation apps represent the most commonly used mental health apps for depression and anxiety [30]. Despite their growing popularity, little is known regarding who in fact benefits from these apps. In the current study, we developed an algorithm to predict the benefit an individual would be expected to experience from a smartphone-based meditation intervention (HMP) relative to an assessment-only control condition. We found evidence that a data-driven model can successfully predict differential response to a meditation app vs. an assessment-only control condition using self-reported baseline demographic and clinical characteristics. Specifically, PAI scores significantly moderated group differences in outcome. Individuals with more negative PAI scores – reflecting relatively stronger meditation app (i.e., HMP) recommendations – had better outcomes if randomly assigned to the meditation app relative to the control condition. As expected given the positive effect of treatment condition on reductions in change in distress [31], the models typically predicted greater benefit from HMP versus the control condition. However, the predicted benefits of HMP were not always large and, in some cases, the PAI model predicted either relatively small between-group differences in outcome ("yellow zone" in **Figure 3**) or even better outcomes in the control condition ("red zone"). The former cases could be interpreted as instances in which the costs of engaging in a multiweek meditation app course (e.g., time investment, delay in engaging with other, more helpful interventions) may not be worth the potential benefits.

Critically, a comparison linear regression model which only included information about baseline levels of repetitive negative thinking performed comparably well to a multivariable machine learning model (in contrast, see [60,61]). Repetitive negative thinking moderated outcome to app-based meditation training relative to the assessment-only control. Importantly, these findings reveal that higher repetitive negative thinking is not simply a general “prognostic” indicator of one’s likelihood of experiencing reductions in distress (e.g., due to regression to the mean or the passage of time). In other words, greater repetitive negative thinking did not predict greater reductions in distress in *both* the meditation app and control condition. Instead, and similar to prior research focused on a different mindfulness app and sample (adolescents with elevated rumination) [51,54], individuals with higher baseline levels of repetitive negative thinking derived greater relative benefit from a meditation app. One question is whether these findings are specific to repetitive negative thinking, or instead may be driven by correlated clinical characteristics, in particular depressive symptoms or distress. Sensitivity analyses revealed that repetitive negative thinking significantly moderated group differences in outcome even when controlling for depressive symptom severity or distress (see Supplement). In summary, these findings indicate that a brief self-report assessment of repetitive negative thinking could inform which individuals are most likely to benefit from app-based meditation training.

As illustrated in **Figure 3**, our predictive model can be readily applied for personalized meditation app recommendations for new individuals. First, the model provides a binary prediction of whether or not an individual is expected to experience greater reductions in distress from the meditation app relative to symptom assessment only (i.e., based on whether PAI scores fall to the left or right of the intersection point [vertical dashed grey line]). Second, the model provides an estimate of the *magnitude* of the expected difference in outcome between the meditation app and control condition. Finally, the model also distinguishes between strengths of recommendations to use the meditation app, demarcated by the green (confident recommendation) and yellow (cautious recommendations) zones of the figure (with boundaries defined by a bootstrapped confidence interval and Johnson-Neyman interval). Collectively, this information could be used to provide individuals with objective metrics about expected outcomes to inform their decision about whether to enroll in a meditation app course. Such information could readily be implemented within mHealth interventions like the HMP. Participants could first complete a brief self-report assessment of repetitive negative thinking and receive feedback on their predicted outcomes *prior* to them deciding to use the app.

While potentially useful in terms of encouraging optimal use of users' time and attention, informing some individuals that engagement with a meditation app may not be beneficial to them is unlikely to be embraced by many intervention developers. However, these models could be readily extended to instances in which one or more mHealth interventions are being compared. Given the thousands of available mental health apps [25], which should be compared? One approach is to focus on the most popular (e.g., most frequently downloaded) mental health apps, which include mindfulness, journaling, CBT and mood tracking apps [29,30]. For example, future studies could develop algorithms for predicting response to various popular mental health apps, which differ substantially in intervention focus (e.g., meditation app vs. CBT-based app vs. mood tracking) [29,62], or even compare a mental health app vs. conventional (in-person) psychotherapy or pharmacotherapy. Such studies could determine, for example, whether we can predict which individuals with depressive symptoms require conventional, face-to-face CBT (or an antidepressant prescription) vs. those who would experience symptom remission from a brief app-based meditation or CBT course. In addition, future studies could compare different versions of a single app. For example, individuals may differ in the extent to which they benefit from different types of meditation (e.g., cultivating focused attention on the breath, open monitoring, or loving-kindness meditations) or different lengths or frequency of guided meditation sessions.

In addition to informing consumer choice, the ability to predict who is most likely to benefit from a particular intervention could inform health care policy and decision making. In contrast to a stepped care model in which treatment intensity is escalated based on response to interventions, predictive models could be used to initially assign patients to the treatment expected to yield the best outcomes for that individual based on their baseline characteristics (i.e., stratified care) [63]. In theory, this latter approach could minimize delay to receiving an effective intervention.

Another important avenue for future research is test the extent to which these findings generalize to other meditation apps (e.g., Headspace and Calm). In many ways, HMP is similar to other meditation apps. It includes training in mindfulness and connection (e.g., loving-kindness, compassion) practices that are also available in popular mindfulness apps like Headspace and Calm. One difference is that HMP includes practices designed specifically to cultivate a healthy sense of self (Insight module) as well as purposes and meaning in life (Purpose module). Inclusion of these practices is derived from the neuroscience-based model of well-being on which HMP is based [36]. Thus, it is more accurate to view HMP as a meditation app that intentionally moves beyond mindfulness to have equal emphasis on other domains of well-being and contemplative practices designed to support these additional domains. Ultimately, additional research is needed to test whether the pattern of findings presented in this study generalize to other meditation/mindfulness apps.

Finally, given the lack of prior research predicting mental health app outcome, research is needed to test the impact of presenting predicted mindfulness app prognosis on patient outcomes. For example, prior to using a mindfulness app, patients could be randomly assigned to receiving their predicted outcomes vs. not receiving this information. Several relevant outcomes could be examined, including: (1) between-group differences in symptom change; (2) the extent to which receiving these predictions influence expectancies of therapeutic benefit; (3) the relation between expectancies and app outcome and (4) the extent to which individuals in fact use the algorithm-recommended intervention vs. disregard the recommendation.

Limitations

The current study has several important limitations. First, although basing models exclusively on self-report data is attractive from an implementation perspective, we may have excluded other patient characteristics which provide important additional predictive information to inform optimal treatment recommendations (e.g., biomarkers, cognitive tasks) [12]. In addition, repetitive negative thinking, which emerged as a predictor of differential response, may be more validly assessed via methods other than conventional, retrospective self-report questionnaires (e.g., repeated, daily ecological momentary assessment [e.g., 51,64]). Relatedly, other relevant variables (e.g., app usage data, motivational variables, involvement in other activities linked to better mental health) could be assessed in future studies. Second, our results emerged within a specific sample (school district employees), which did not have adequate representation of males, BIPOC or lower-income individuals. The sample is representative of Wisconsin in terms of race (83% of Wisconsinites are White) but includes a higher proportion of females. However, the gender difference in our sample is not surprising given that females are more likely than males to: (1) be employed as teachers [65] and (2) experience, and seek treatment for, depressive and anxiety symptoms [66]. Third, we were unable to conduct external validation by evaluating performance in an entirely new sample (e.g., from another RCT). Fourth, as is common in digital therapies [67], a sizable subset of participants used the app for relatively few days. However, results remained significant when restricting our analyses to subsets of participants who used the app for a longer period (see Supplement). Fifth, we did not include an active comparison condition. Our assessment-only control condition was not designed to control for placebo-related processes [68]. The methods demonstrated here may ultimately be most relevant in helping patients and clinicians decide between competing interventions that are intended to be therapeutic.

Conclusions and Future Directions

The current study demonstrates the potential utility of data-driven approaches to informing personalized meditation app recommendations. A natural extension of this study would be to conduct a prospective test of our algorithm using a doubly randomized design. For example, participants could be randomized to either: (1) random treatment assignment (i.e., Treatment A or Treatment B) or to (2) be assigned to their algorithm-indicated treatment. To the extent that patient outcomes are significantly (and clinically meaningfully) better in the latter condition, results would support the clinical benefits of algorithm-informed treatment recommendations (for a recent example of a similar design testing predictive matching of patients to therapists, see [69]). In addition to comparing treatment packages, this design could be readily used to evaluate other customizable elements of HMP or other mHealth interventions. This may include assignment to receive various components or ordering of components within HMP, assignment to HMP or an alternative commonly used mHealth intervention (e.g., CBT, behavioral activation, journaling, or mood tracking apps), or assignment to varying treatment intensities (e.g., meditation practice frequency).

Other potentially fruitful future directions include evaluating a broader set of patient characteristics previously shown or hypothesized to predict likelihood of response to different interventions [5]. In addition, prediction models could be developed using data drawn from large naturalistic datasets evaluating mHealth interventions, as has been done for in-person psychotherapy and pharmacotherapy [60,70–72]. In addition to testing the utility of these models in “real-world” settings, naturalistic settings often provide large datasets relative to RCTs and thus can increase statistical power [28]. Ultimately, these approaches may gradually help supplant our reliance on trial-and-error for treatment selection with empirically supported, data-driven algorithms to objectively communicate expected benefits to individuals, allowing them to make well-informed decisions about which interventions are best for their needs.

Acknowledgments

This research was supported by NCCIH Grant K23AT010879 (SBG), NIMH Grant R01MH43454 (RJD), Chan Zuckerberg Initiative Grant 2020-218037 (RJD), a National Academy of Education/Spencer Postdoctoral Fellowship (MJH), and with funding from the Wisconsin Center for Education Research (SBG). The first author (CAW) was partially supported by NIMH R01MH116969, NCCIH R01AT011002, the Tommy Fuss Fund and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation.

Conflicts of Interest

RJD is the founder, president, and serves on the board of directors for the nonprofit organization, Healthy Minds Innovations, Inc. MJH has been a paid consultant at Healthy Minds Innovations, Inc. for work unrelated to this research.

References

1. Collins FS, Varmus H. A new initiative on precision medicine. New England journal of medicine Mass Medical Soc; 2015;372(9):793–795. [doi: 10.1056/NEJMp1500523]

2. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science American Association for the Advancement of Science; 2004;304(5676):1497–1500. [doi: 10.1126/science.1099314]

3. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. The lancet oncology Elsevier; 2012;13(3):239–246. [doi: 10.1016/S1470-2045(11)70393-X]

4. Schwaederle M, Zhao M, Lee JJ, Eggermont AM, Schilsky RL, Mendelsohn J, Lazar V, Kurzrock R. Impact of precision medicine in diverse cancers: a meta-analysis of phase II clinical trials. Journal of clinical oncology American Society of Clinical Oncology; 2015;33(32):3817. [doi: 10.1200/JCO.2015.61.5997]

5. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Ebert DD, de Jonge P, Nierenberg AA, Rosellini AJ, Sampson NA, Schoevers RA, Wilcox MA, Zaslavsky AM. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. Epidemiol Psychiatr Sci 2017 Feb;26(1):22–36. PMID:26810628 [doi: 10.1017/S2045796016000020]

6. Norcross JC, Wampold BE. A new therapy for each patient: Evidence-based relationships and responsiveness. Journal of Clinical Psychology Wiley Online Library; 2018;74(11):1889–1906. [doi: 10.1002/jclp.22678]

7. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. 2nd edition. New York, NY: Springer; 2017. ISBN:978-0-387-84857-0

8. Webb CA, Cohen ZD. Progress towards clinically informative data-driven decision support tools in psychotherapy. The Lancet Digital Health Elsevier; 2021;3(4):e207–e208. [doi: 10.1016/S2589-7500(21)00042-X]

9. Boehmke B, Greenwell BM. Hands-On Machine Learning with R. 1st edition. Boca Raton: Chapman and Hall/CRC; 2019. ISBN:978-1-138-49568-5

10. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annual review of clinical psychology Annual Reviews; 2018;14:91–118. [doi: 10.1146/annurev-clinpsy-032816-045037]

11. Aafjes-van Doorn K, Kamsteeg C, Bate J, Aafjes M. A scoping review of machine learning in psychotherapy research. Psychotherapy Research Taylor & Francis; 2021;31(1):92–116. [doi: 10.1080/10503307.2020.1808729]

12. Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, Cohen Z, Belgrave D, DeRubeis R, Iniesta R. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry Wiley Online Library; 2021;20(2):154–170. [doi: 10.1002/wps.20882]

13. Cohen ZD, DeRubeis RJ. Treatment selection in depression. Annual Review of Clinical Psychology 2018;14(1):209–236. PMID:29494258 [doi: 10.1146/annurev-clinpsy-050817-084746]

14. Delgadillo J, Lutz W. A development pathway towards precision mental health care. JAMA psychiatry American Medical Association; 2020;77(9):889–890. [doi: 10.1001/jamapsychiatry.2020.1048]

15. DeRubeis RJ. The history, current status, and possible future of precision mental health. Behaviour Research and Therapy 2019 Dec 1;123:103506. [doi: 10.1016/j.brat.2019.103506]

16. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment Recommendations. A Demonstration. PLOS ONE 2014 Jan 8;9(1):e83875. [doi: 10.1371/journal.pone.0083875]

17. Huibers MJH, Cohen ZD, Lemmens LHJM, Arntz A, Peeters FPML, Cuijpers P, DeRubeis RJ. Predicting Optimal Outcomes in Cognitive Therapy or Interpersonal Psychotherapy for Depressed Individuals Using the Personalized Advantage Index Approach. PLOS ONE 2015 Nov 10;10(11):e0140771. [doi: 10.1371/journal.pone.0140771]

18. Cohen ZD, Kim TT, Van HL, Dekker JJM, Driessen E. A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive–behavioral versus psychodynamic therapy for mild to moderate adult depression. Psychotherapy Research 2019 Jan 11;0(0):1–14. PMID:30632922 [doi: 10.1080/10503307.2018.1563312]

19. Webb CA, Trivedi MH, Cohen ZD, Dillon DG, Fournier JC, Goer F, Fava M, McGrath PJ, Weissman M, Parsey R, Adams P, Trombello JM, Cooper C, Deldin P, Oquendo MA, McInnis MG, Huys Q, Bruder G, Kurian BT, Jha M, DeRubeis RJ, Pizzagalli DA. Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. Psychological Medicine 2018 Jul;1–10. [doi: 10.1017/S0033291718001708]

20. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. World Psychiatry Wiley Online Library; 2017;16(1):90–99. [doi: 10.1002/wps.20388]

21. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bulletin of the World Health Organization 2004 Nov;82(11):858–866. [doi: 10.1590/S0042-96862004001100011]

22. Andersson G, Cuijpers P. Internet-Based and Other Computerized Psychological Treatments for Adult Depression: A Meta-Analysis. Cognitive Behaviour Therapy 2009 Dec;38(4):196–205. [doi: 10.1080/16506070903318960] [doi: 10.1080/16506070903318960]

23. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. World Psychiatry Wiley Online Library; 2019;18(3):325–336. [doi: 10.1002/wps.20673]

24. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? Jama American Medical Association; 2013;310(22):2395–2396. [doi: 10.1001/jama.2013.281078]

25. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, Haim A, Hsin H, Hollis C, Lewis S. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry World Psychiatric Association; 2019;18(1):97. [doi: 10.1002/wps.20592]

26. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, Tschiatschek S, Nori A, Thieme A, Richards D. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA network open American Medical Association; 2020;3(7):e2010791–e2010791. [doi: 10.1001/jamanetworkopen.2020.10791]

27. Eysenbach G. The law of attrition. Journal of medical Internet research JMIR Publications Inc., Toronto, Canada; 2005;7(1):e11. [doi: 10.2196/jmir.7.1.e11]

28. Luedtke A, Sadikova E, Kessler RC. Sample Size Requirements for Multivariate Models to Predict Between-Patient Differences in Best Treatments of Major Depressive Disorder. Clinical Psychological Science 2019 May 1;7(3):445–461. [doi: 10.1177/2167702618815466]

29. Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing Mental Health Apps Marketplaces with Objective Metrics from 29,190 Data Points from 278 Apps. Acta Psychiatrica Scandinavica Wiley Online Library; 2021; [doi: 10.1111/acps.13306]

30. Wasil AR, Gillespie S, Shingleton R, Wilks CR, Weisz JR. Examining the reach of smartphone apps for depression and anxiety. AJP American Psychiatric Publishing; 2020 May 1;177(5):464–465. [doi: 10.1176/appi.ajp.2019.19090905]

31. Hirshberg MJ, Frye C, Dahl CJ, Riordan K, Vack N, Sachs J, Goldman RI, Davidson RJ, Goldberg S. A randomized controlled trial of a smartphone-based well-being training in

public school system employees during the COVID-19 pandemic. *Journal of Educational Psychology.* Advance online publication. https://doi.org/10.1037/edu0000739

32. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA network open American Medical Association; 2020;3(9):e2019686–e2019686. [doi: 10.1001/jamanetworkopen.2020.19686]

33. Markowitz AJ, Bassok D, Bellows L, Doromal JB, Michie M, Smith A. Understanding Early Educators' Responses to the Coronavirus Pandemic: Lessons from Large-Scale Surveys in Louisiana and Virginia. 2020 APPAM Fall Research Conference APPAM; 2020.

34. Markowitz AJ, Ehrlich SB, Gassman-Pines A. Child Care in the Time of COVID-19: New Evidence from Child Care Business Owners, Early Educators, and Parents to Inform Recovery Efforts. 2020 APPAM Fall Research Conference APPAM; 2020.

35. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, Group PC. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment Sage Publications Sage CA: Los Angeles, CA; 2011;18(3):263–283. [doi: 10.1177/1073191111411667]

36. Dahl CJ, Wilson-Mendenhall CD, Davidson RJ. The plasticity of well-being: A training-based framework for the cultivation of human flourishing. PNAS 2020 Dec 22;117(51):32197–32206. [doi: 10.1073/pnas.2014859117]

37. Goldberg SB, Imhoff-Smith T, Bolt DM, Wilson-Mendenhall CD, Dahl CJ, Davidson RJ, Rosenkranz MA. Testing the Efficacy of a Multicomponent, Self-Guided, Smartphone-Based Meditation App: Three-Armed Randomized Controlled Trial. JMIR Mental Health 2020 Nov 27;7(11):e23825. [doi: 10.2196/23825]

38. Cohen S. Perceived stress in a probability sample of the United States. Sage Publications, Inc; 1988;

39. Roberti JW, Harrington LN, Storch EA. Further psychometric support for the 10-item version of the perceived stress scale. Journal of College Counseling Wiley Online Library; 2006;9(2):135–147. [doi: 10.1002/j.2161-1882.2006.tb00100.x]

40. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, Riley W, Cella D. Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. Journal of clinical epidemiology Elsevier; 2016;73:119–127. [doi: 10.1016/j.jclinepi.2015.08.036]

41. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. Journal of behavior therapy and experimental psychiatry Elsevier; 2011;42(2):225–232. [doi: 10.1016/j.jbtep.2010.12.003]

42. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychotherapy and psychosomatics Karger Publishers; 2015;84(3):167–176. [doi: 10.1159/000376585]

43. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, Walsh E, Duggan D, Williams JMG. Construct Validity of the Five Facet Mindfulness Questionnaire in Meditating and Nonmeditating Samples. Assessment 2008 Sep 1;15(3):329–342. [doi: 10.1177/1073191107313003]

44. Cyranowski JM, Zill N, Bode R, Butt Z, Kelly MA, Pilkonis PA, Salsman JM, Cella D. Assessing social support, companionship, and distress: National Institute of Health (NIH) Toolbox Adult Social Relationship Scales. Health Psychology American Psychological Association; 2013;32(3):293. [doi: 10.1037/a0028586]

45. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the self-compassion scale. Clinical psychology & psychotherapy Wiley Online Library; 2011;18(3):250–255. [doi: 10.1002/cpp.702]

46. Forman EM, Herbert JD, Juarascio AS, Yeomans PD, Zebell JA, Goetter EM, Moitra E. The Drexel defusion scale: A new measure of experiential distancing. Journal of Contextual Behavioral Science Elsevier; 2012;1(1–2):55–65. [doi: 10.1016/j.jcbs.2012.09.001]

47. Steger MF, Frazier P, Oishi S, Kaler M. The meaning in life questionnaire: Assessing the presence of and search for meaning in life. Journal of counseling psychology American Psychological Association; 2006;53(1):80. [doi: 10.1037/0022-0167.53.1.80]

48. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 2012 Jan 1;28(1):112–118. [doi: 10.1093/bioinformatics/btr597]

49. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine American College of Physicians; 2015;162(1):W1–W73. [doi: 10.7326/M14-0698]

50. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. NeuroImage 2017 Jan 15;145:166–179. [doi: 10.1016/j.neuroimage.2016.10.038]

51. Webb CA, Swords CM, Murray LM, Hilt LM. App-based Mindfulness Training for Adolescent Rumination: Predictors of Immediate and Cumulative Benefit. Mindfulness 2021 Jan 15;12:2498–2509. [doi: https://doi.org/10.1007/s12671-021-01719-0]

52. Wetherill RR, Rao H, Hager N, Wang J, Franklin TR, Fan Y. Classifying and Characterizing Nicotine Use Disorder with High Accuracy Using Machine Learning and Resting-State fMRI. Addict Biol 2019 Jul;24(4):811–821. PMID:29949234 [doi: 10.1111/adb.12644]

53. Schwartz B, Cohen ZD, Rubel JA, Zimmermann D, Wittmann WW, Lutz W. Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. Psychotherapy Research Taylor & Francis; 2020;1–19. [doi: 10.1080/10503307.2020.1769219]

54. Webb CA, Swords CM, Lawrence H, Hilt LM. Which adolescents are well-suited to app-based mindfulness training? A randomized clinical trial and data-driven approach for personalized recommendations [Internet]. Journal of Consulting and Clinical Psychology. 2022 [cited 2021 Jul 21]. [doi: 10.31234/osf.io/7gbjd]

55. Team. R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing [Internet] 2021; Available from: https://www.R-project.org/

56. Watkins ER, Nolen-Hoeksema S. A habit-goal framework of depressive rumination. J Abnorm Psychol 2014 Feb;123(1):24–34. PMID:24661156 [doi: 10.1037/a0035540]

57. Watkins ER, Roberts H. Reflecting on rumination: Consequences, causes, mechanisms and treatment of rumination. Behaviour Research and Therapy 2020 Apr 1;127:103573. [doi: 10.1016/j.brat.2020.103573]

58. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-28.; 2021.

59. Hayes AF, Matthes J. Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. Behavior research methods Springer; 2009;41(3):924–936. [doi: 10.3758/BRM.41.3.924]

60. Webb CA, Cohen ZD, Beard C, Forgeard M, Peckham AD, Björgvinsson T. Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. Journal of Consulting and Clinical Psychology US: American Psychological Association; 2020;88(1):25–38. [doi: 10.1037/ccp0000451]

61. Buckman JE, Cohen ZD, O'Driscoll C, Fried EI, Saunders R, Ambler G, DeRubeis RJ, Gilbody S, Hollon SD, Kendrick T. Predicting prognosis for adults with depression using individual symptom data: a comparison of modelling approaches. Psychological Medicine Cambridge University Press; 2021;1–11. [doi: 10.1017/S0033291721001616]

62. Flett JA, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile Mindfulness Meditation: a Randomised Controlled Trial of the Effect of Two Popular Apps on Mental Health. Mindfulness 2019;10(5):863–876. [doi: 10.1007/s12671-018-1050-9]

63. Lipton RB, Stewart WF, Stone AM, Láinez MJ, Sawyer JP. Stratified care vs step care strategies for migraine: the Disability in Strategies of Care (DISC) Study: a randomized trial. Jama American Medical Association; 2000;284(20):2599–2605. [doi: 10.1001/jama.284.20.2599]

64. Webb CA, Israel ES, Belleau E, Appleman L, Forbes EE, Pizzagalli DA. Mind-Wandering in Adolescents Predicts Worse Affect and Is Linked to Aberrant Default Mode Network–Salience Network Connectivity. Journal of the American Academy of Child & Adolescent Psychiatry 2021 Mar 1;60(3):377–387. [doi: 10.1016/j.jaac.2020.03.010]

65. National Center for Education Statistics. Characteristics of Public School Teachers [Internet]. [cited 2021 Oct 6]. Available from: https://nces.ed.gov/programs/coe/indicator/clr

66. Eaton NR, Keyes KM, Krueger RF, Balsis S, Skodol AE, Markon KE, Grant BF, Hasin DS. An Invariant Dimensional Liability Model of Gender Differences in Mental Disorder Prevalence: Evidence from a National Sample. J Abnorm Psychol 2012 Feb;121(1):282–288. PMID:21842958 [doi: 10.1037/a0024780]

67. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evidence-based mental health Royal College of Psychiatrists; 2018;21(3):116–119. [doi: 0.1136/eb-2018-102891]

68. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. The Lancet Psychiatry Elsevier; 2016;3(2):100–102. [doi: 10.1016/S2215-0366(15)00565-9]

69. Constantino MJ, Boswell JF, Coyne AE, Swales TP, Kraus DR. Effect of Matching Therapists to Patients vs Assignment as Usual on Adult Psychotherapy Outcomes: A Randomized Clinical Trial. JAMA psychiatry 2021; [doi: 10.1001/jamapsychiatry.2021.1221]

70. Bone C, Simmonds-Buckley M, Thwaites R, Sandford D, Merzhvynska M, Rubel J, Deisenhofer A-K, Lutz W, Delgadillo J. Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data. The Lancet Digital Health Elsevier; 2021;3(4):e231–e240. [doi: 10.1016/S2589-7500(21)00018-2]

71. Delgadillo J, Appleby S, Booth S, Burnett G, Carey A, Edmeade L, Green S, Griffin P, Johnson E, Jones R. The Leeds Risk Index: Field-test of a stratified psychological treatment selection algorithm. Psychotherapy and psychosomatics Karger Publishers; 2020; [doi: 10.1159/000505193]

72. Webb CA, Forgeard M, Israel ES, Lovell-Smith N, Beard C, Björgvinsson T. Personalized prescriptions of therapeutic skills from patient characteristics: An ecological momentary assessment approach. Journal of Consulting and Clinical Psychology US: American Psychological Association; 2021;No Pagination Specified-No Pagination Specified. [doi: 10.1037/ccp0000555]

Table 1

Descriptive statistics for HMP and assessment-only control at baseline

	HMP			Control			
Variable	**n**	**Mean / %**	**SD**	**n**	**Mean / %**	**SD**	**p**
Age	344	42.47	11.06	318	42.7	10.23	.778
Female gender	344	0.87	0.34	318	0.88	0.33	.752
Non-Hispanic White	344	0.88	0.32	318	0.84	0.36	.127
Married	344	0.71	0.46	318	0.68	0.47	.450
College	343	0.90	0.30	316	0.89	0.31	.718
Income $50k or less	344	0.16	0.37	318	0.17	0.38	.727
Income $50-100k	344	0.41	0.49	318	0.41	0.49	.912
Income $100-150k	344	0.30	0.46	318	0.30	0.46	.990
Income $150k+	344	0.12	0.32	318	0.12	0.32	.998
PROMIS Depression	342	55.37	6.20	315	55.47	6.43	.850
PROMIS Anxiety	342	59.83	6.95	315	60.00	7.11	.754
PSS Stress	342	2.89	0.56	315	2.87	0.60	.687
Distress (composite)	342	0.00	0.88	315	0.00	0.91	.971
PTQ Repetitive negative thinking	342	29.89	10.43	315	29.62	11.29	.755
FFMQ Awareness	342	24.80	5.93	315	24.56	6.12	.615
NIHTL Loneliness	342	2.53	0.77	315	2.58	0.77	.445
DDS Defusion	342	24.83	7.89	315	24.50	8.16	.601
MLQ Presence	342	26.20	5.44	315	25.81	5.46	.361
MLQ Search for meaning	342	21.63	6.61	315	22.09	6.79	.381
WHO Well-being	341	12.76	4.71	315	12.47	4.33	.420
SCS Self-compassion	342	2.98	0.69	315	2.93	0.70	.374

Note: PROMIS = Patient-Reported Outcomes Information System; PSS = Perceived Stress Scale; Distress = composite of PROMIS Depression, PROMIS Anxiety, and PSS; PTQ = Perseverative Thinking Questionnaire; FFMQ = Five Facet Mindfulness

Questionnaire; NIHTL = National Institutes of Health Toolbox Loneliness; DDS = Drexel Defusion Scale; MLQ = Meaning in Life Questionnaire; WHO = World Health Organization; SCS = Self-Compassion Scale; *p* = *p*-value from independent samples t-test comparing groups at baseline. Table values include means with the exception of dichotomous variables where the mean column includes percentages (female gender, non-Hispanic White race/ethnicity, married, college education, income brackets). Sample size varied slightly by measure as participants were not required to complete every measure (e.g., if they were uncomfortable responding to items on a given measure or reporting on a given demographic characteristic).

Table 2

Baseline variables retained in elastic net models predicting outcome for each condition

	HMP Model	Control Model
Predictors	**Coefficient**	**Coefficient**
Age		
Gender		
Race		
Marital status		-0.006
Income		
PROMIS Depression	-0.012	-0.005
PROMIS Anxiety		-0.007
PSS Stress	-0.003	-0.006
Distress (composite)	-0.011	-0.008
PTQ Repetitive negative thinking		0.012
FFMQ Awareness		
NIHTL Loneliness		-0.002
DDS Defusion		-0.011
MLQ Presence		-0.008
MLQ Search for meaning		
WHO Well-being		
SCS Self-compassion		-0.002

Note. PROMIS = Patient-Reported Outcomes Information System; PSS = Perceived Stress Scale; Distress = composite of PROMIS Depression, PROMIS Anxiety, and PSS; PTQ = Perseverative Thinking Questionnaire; FFMQ = Five Facet Mindfulness Questionnaire; NIHTL = National Institutes of Health Toolbox Loneliness; DDS = Drexel Defusion Scale; MLQ = Meaning in Life Questionnaire; WHO = World Health Organization; SCS = Self-Compassion Scale. The larger set of baseline predictors retained in the ENR model applied to the control participants relative to the HMP group is due to the fact that the best fitting model in the former group had a lower alpha value (i.e., closer to ridge than lasso regression) relative to the HMP group. Negative parameter estimates indicate that higher scores on the predictor variable are associated with better outcome (i.e., reductions in distress).

Figure Captions

Figure 1. Group x Personalized Advantage Index (PAI) interaction. As PAI scores decrease (i.e., reflecting relatively stronger recommendations for the Healthy Minds Program [HMP] app) group differences in observed outcome increase, favoring HMP.

Figure 2. Group x Personalized Advantage Index (PAI) interaction for the comparison model (i.e., linear regression with baseline repetitive negative thinking (PTQ) scores as the sole predictor). As PAI scores decrease (i.e., reflecting relatively stronger recommendations for the Healthy Minds Program [HMP] app) group differences in observed outcome increase, favoring HMP.

Figure 3. Plot of the relationship between Personalized Advantage Index (PAI) scores and outcome for each condition to inform personalized recommendations. The dashed vertical grey line indicates the point at which the two regression lines intersect (left margin of a bootstrapped 95% confidence interval is shown with a dashed vertical red line). The solid vertical grey line (adjacent to the red line) is derived from the Johnson-Neyman technique and represents the value of the moderator (PAI) at which between-group differences in outcome become statistically significant. See detailed description in text, with an example for personalized HMP recommendation.

Figure 1

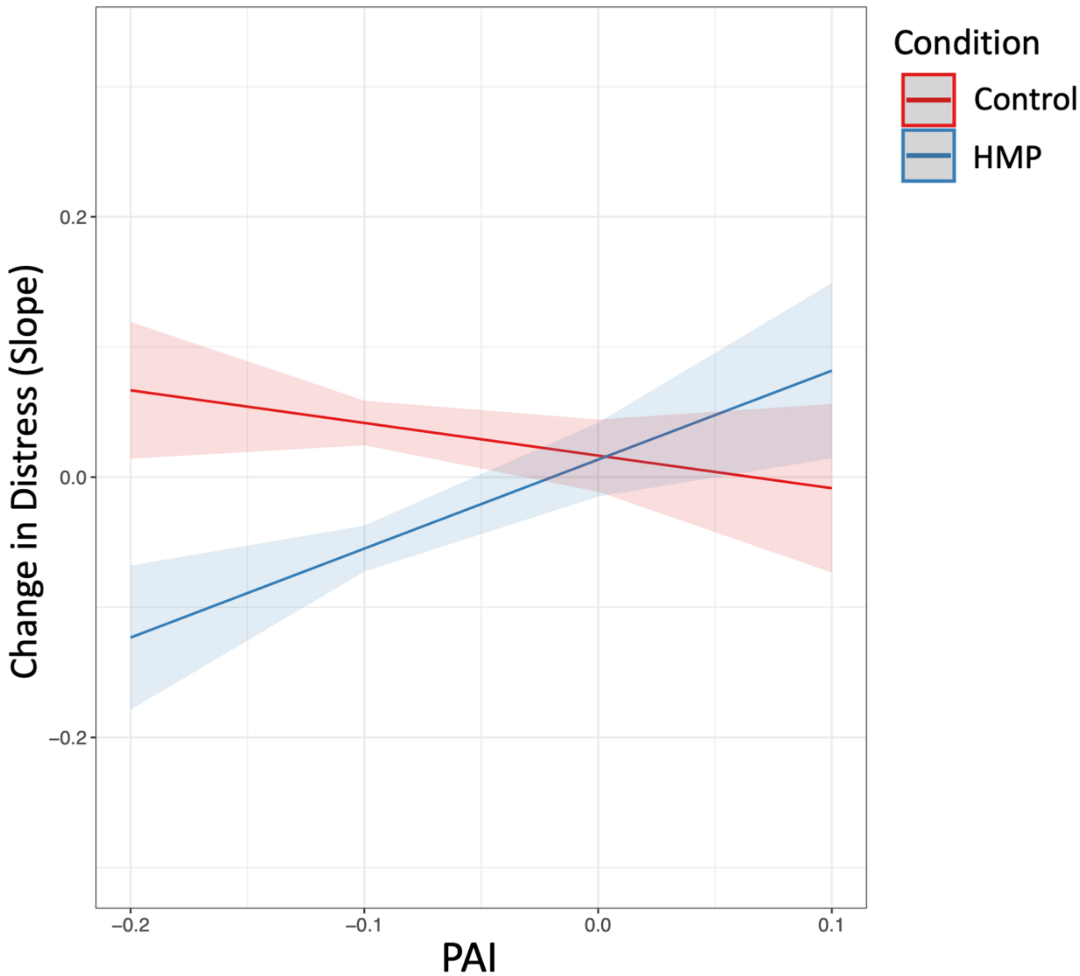

Condition
Control
HMP
Change in Distress (Slope)
0.2
0.0
−0.2
−0.2
−0.1
0.0
0.1
PAI

Figure 2

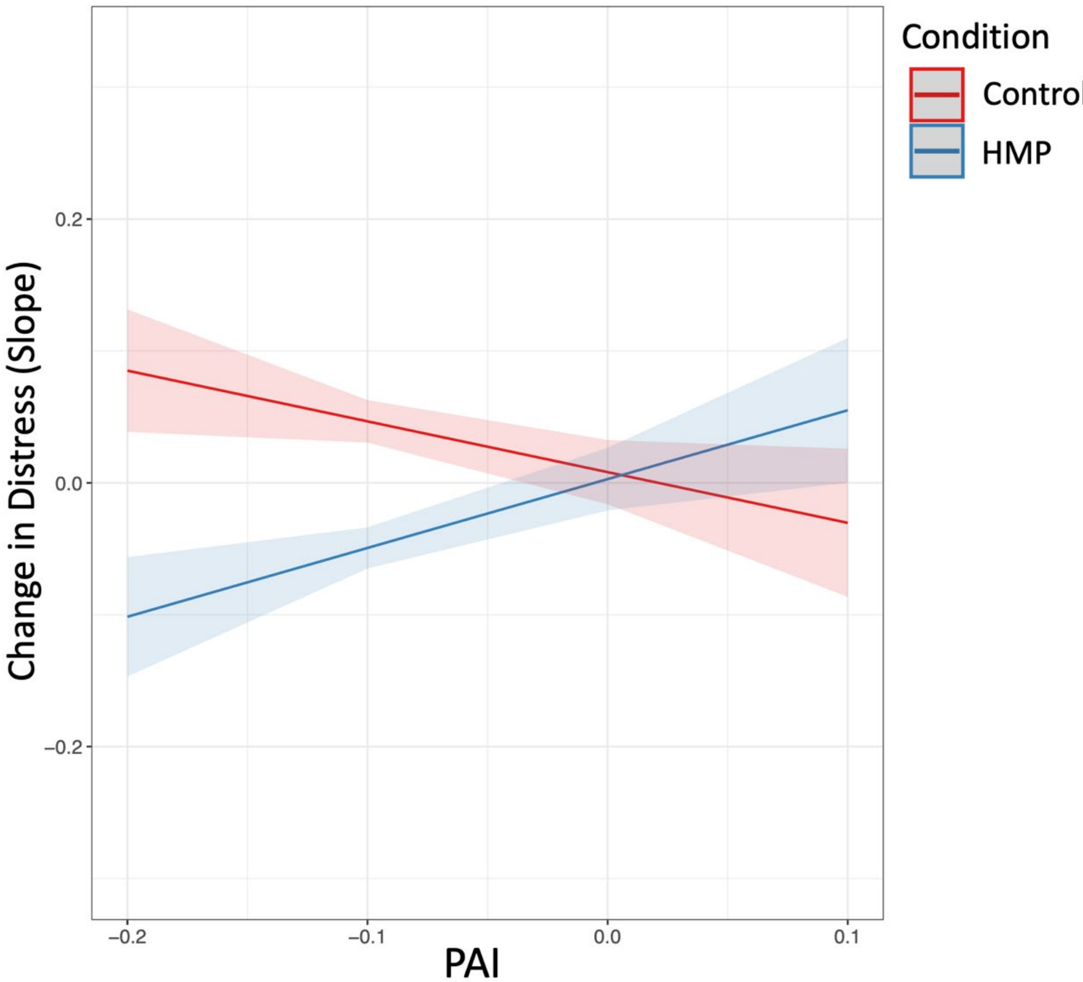
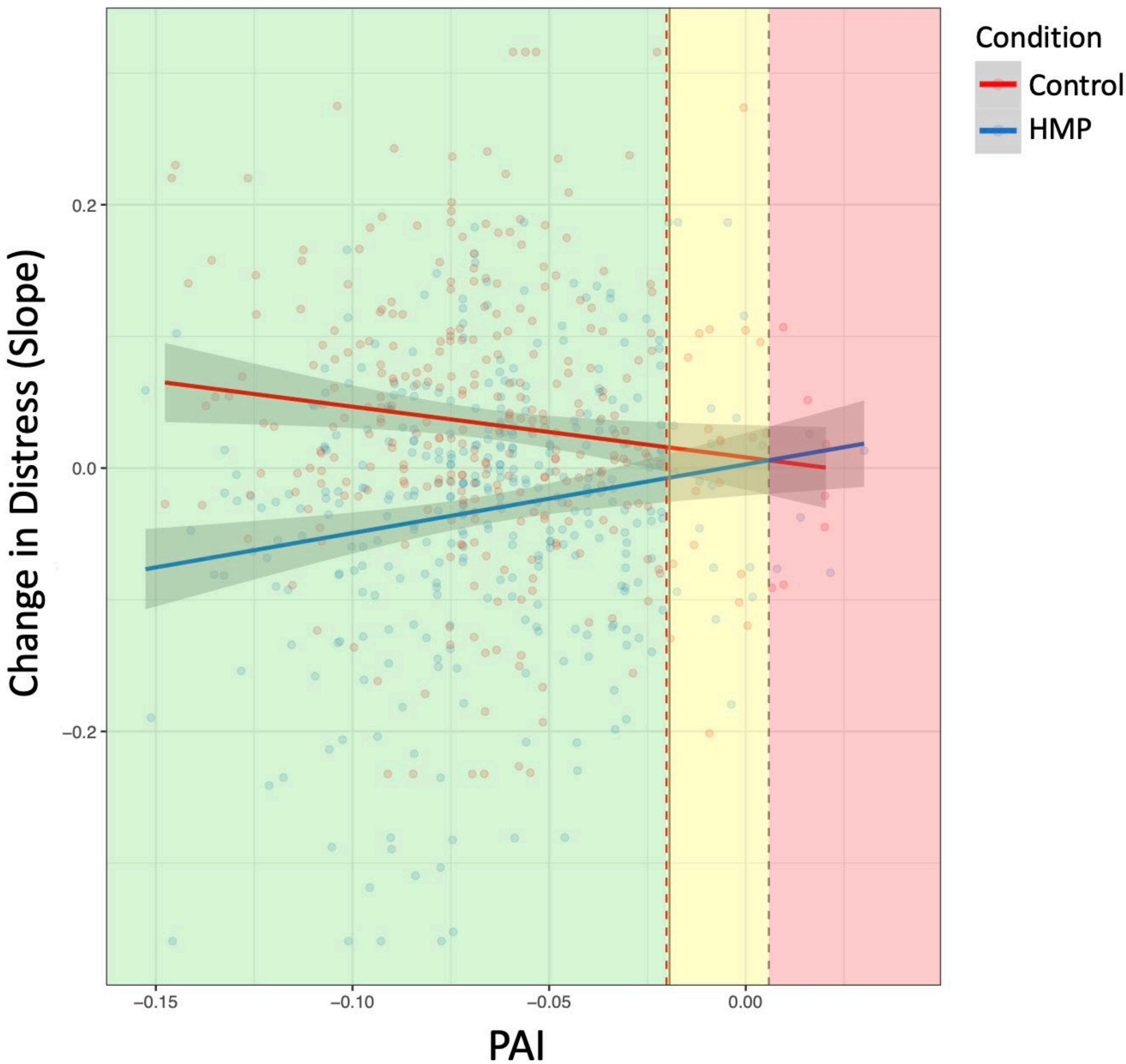

Condition
Control
HMP
Change in Distress (Slope)
0.2
0.0
−0.2
−0.2
−0.1
0.0
0.1
PAI

Figure 3


Condition
Control
HMP
Change in Distress (Slope)
0.2
0.0
−0.2
−0.15
−0.10
−0.05
0.00
PAI

Supplementary Online Content

Personalized prediction of response to smartphone-delivered meditation training

Supplemental Methods

Healthy Minds Program (HMP)

The HMP includes contemplative practices designed to build skills supportive of four pillars of well-being: awareness, connection, insight, and purpose [1,2]. Briefly, awareness includes skills in attention regulation and meta-awareness; connection involves intra- and interpersonal relational skills including gratitude, kindness, and compassion; insight is structured around an accurate understanding of how beliefs regarding identity and self shape experience; and purpose involves clarifying values and applying them in daily life activities. Participants were encouraged to engage with content from each of the four modules for approximately one week (i.e., 4 weeks total). Content included didactic instruction as well as guided meditation practices. For the guided practices, participants could select the length of practice from 5 to 30 minutes. The HMP app was used for a mean of 10.9 days (SD = 9) over the 4-week trial. Within digital therapies, attrition and low uptake is a common challenge [3]. In the present study, 73% of the sample used the app for 2 or more days (69% for 3 or more days, 67% for 4 or more days). For additional trial and sample details, see [4].

Primary Outcome

The prespecified primary outcome in the parent RCT was psychological distress which was a composite of the computer-adaptive versions of the PROMIS Anxiety and PROMIS Depression measures [5] and the 10-item Perceived Stress Scale (PSS) [6]. All three are widely used measures with established reliability and validity [7,8]. Internal consistency of the PSS was adequate in the current sample (a = .86). A composite was calculated by averaging across standardized (i.e., z-scored) versions of the three measures (standardized using baseline means and standard deviation [SD])(Given that we used the computer adaptive PROMIS, Cronbach's alpha cannot be computed for this measure).

Post-Imputation Procedures

Following the multiple imputation, and prior to conducting the prediction analysis, continuous variables were z-standardized (mean = 0, SD = 1) and categorical variables (i.e., marital status, gender and race) dummy coded (-0.5 and 0.5). Given the small percentage of non-White participants in this sample (86% non-Hispanic White), race was coded as White or non-White. To reduce the influence of outliers, we winsorized extreme values (Winsorize function in the DescTools R package) by setting values below the 1st percentile and above the 99th percentile to the 1st percentile and 99th percentile values, respectively [9].

Generating predicted outcomes

To predict outcomes, two prognostic models were developed, one for participants who received HMP and one for those who received the assessment-only control condition. To minimize overfitting which can occur with traditional *k*-fold cross-validation (CV), a nested CV procedure was used for each of these prognostic models (i.e., incorporating an outer and inner CV loop [10–13]). For the nested CV, we first split the data into 10 folds (10 training/test sets), representing the outer CV loop. For *each* of the latter outer training sets, the above set of predictor variables were submitted to 10-fold CV (i.e., the "inner" CV loop) elastic net regularized regression (ENR; glmnet package) to generate predictions of outcome (repeated 100 times to generate stable estimates). Specifically, each of the outer training samples were split into 10 equal-sized samples and predicted outcomes for each of the held out 1/10 of the training sample were generated from an ENR model developed in the other 9/10ths of the data. ENR's alpha (which controls the balance between ridge regression [alpha = 0] and LASSO [alpha = 1]) and lambda (which controls the extent to which predictor coefficients are shrunk) parameters were tuned via the CARET package's tuneLength parameter which was set to 20 resulting in 400 combinations of alpha and lambda values. The combination of alpha and lambda that minimized root mean squared error (RMSE; estimated with the inner CV) was selected, and a final model was fit on the entire outer training set and used to predict outcomes for the participants in the outer test set. Importantly, the nested CV procedure ensures that predicted HMP and control condition outcomes for all participants were generated from ENR models that were constructed without the use of their own data.

Supplemental Results

Sample Demographics. Consistent with the demographics of Wisconsin school district employees, the sample was predominantly female (86.9%) and non-Hispanic White (86.1%). Most were married (69.5%) and had completed a college degree (89.2%). The most common income bracket was US$50,000-US$100,000 (40.9%), followed by US$100,000-US$150,000 (30.4%).

Sensitivity Analyses. Several alternative baseline comparison models were run for the purpose of sensitivity analyses. First, in the main text we report a significant Group x PAI interaction in predicting symptom change. We re-ran this analysis controlling for baseline distress which yielded the same pattern of findings ($t(657) = 3.07$, $P = .003$; adjusted $r^2 = .14$). Second, given the association between repetitive negative thinking and depressive symptoms [14,15], we also conducted a sensitivity analysis controlling for baseline levels of depressive symptoms, which yielded the same pattern of findings (Group x PAI interaction, ($t(657) = 3.67$, *P* < .001; adjusted r^2 = .14). Third, we reran the comparison PAI models substituting baseline distress scores (i.e., pre-intervention scores on the outcome measure) as the sole predictor in the model for baseline repetitive negative thinking scores, which yielded the same significant interactions, though slightly attenuated (Group x PAI interaction, ($t(658) = 2.20$, $P = .028$; adjusted r^2 = .13). Fourth, we tested whether baseline repetitive negative thinking scores moderated group differences in outcome after controlling for a baseline depressive symptom severity x group interaction. The baseline repetitive negative thinking x group interaction remained significant ($t(656) = -2.61$, $P = .009$), whereas the baseline depression x group interaction was not significant ($t(656) = -1.28$, $P = .200$). Similarly, the baseline repetitive negative thinking x group interaction remained significant ($t(656) = -2.19$, $P = .029$) when controlling for a baseline distress x group interaction ($t(656) = -0.99$, $P = .320$). Finally, consistent with elevated attrition and low uptake from digital therapies,[3] 73% of the sample used the app for 2 or more days (69% for 3 or more days, 67% for 4 or more days). Consistent with the primary outcome report [4], we included all subjects in our intent-to-treat analyses. However, we re-ran the primary group x PAI interaction on a dataset restricted to those whose used the app 2 or more days. The interaction remained significant ($t(559) = 3.20$, $P = .001$). Similarly, results remained significant when restricting our results to subsamples using the app for 3 or more days ($t(544) = 3.11$, $P = .002$), for at least one day per week (i.e., 4 or more days total)($t(539) = 3.14$, $P = .002$) or at least once daily (i.e., 28 or more times)($t(316) = 3.59$, $P < .001$).

Supplementary References

1. Dahl CJ, Wilson-Mendenhall CD, Davidson RJ. The plasticity of well-being: A training-based framework for the cultivation of human flourishing. Proc Natl Acad Sci 2020 Dec 22;117(51):32197–32206.

2. Goldberg SB, Imhoff-Smith T, Bolt DM, Wilson-Mendenhall CD, Dahl CJ, Davidson RJ, Rosenkranz MA. Testing the Efficacy of a Multicomponent, Self-Guided, Smartphone-Based Meditation App: Three-Armed Randomized Controlled Trial. JMIR Ment Health 2020 Nov 27;7(11):e23825. [doi: 10.2196/23825]

3. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health Royal College of Psychiatrists; 2018;21(3):116–119.

4. Hirshberg MJ, Frye C, Dahl CJ, Riordan K, Vack N, Sachs J, Goldman RI, Davidson RJ, Goldberg S. A Randomized Controlled Trial of a Smartphone-Based Well-Being Training in Public School System Employees During the COVID-19 Pandemic [Internet]. PsyArXiv; 2021 [cited 2021 Jun 26]. [doi: 10.31234/osf.io/hrvmu]

5. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, Group PC. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment Sage Publications Sage CA: Los Angeles, CA; 2011;18(3):263–283.

6. Cohen S. Perceived stress in a probability sample of the United States. Sage Publications, Inc; 1988;

7. Roberti JW, Harrington LN, Storch EA. Further psychometric support for the 10-item version of the perceived stress scale. J Coll Couns Wiley Online Library; 2006;9(2):135–147.

8. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, Riley W, Cella D. Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. J Clin Epidemiol Elsevier; 2016;73:119–127.

9. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. 2nd ed. 2019 edition. Cham, Switzerland: Springer; 2019. ISBN:978-3-030-16398-3

10. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med American College of Physicians; 2015;162(1):W1–W73.

11. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. NeuroImage 2017 Jan 15;145:166–179. [doi: 10.1016/j.neuroimage.2016.10.038]

12. Webb CA, Swords CM, Murray LM, Hilt LM. App-based Mindfulness Training for Adolescent Rumination: Predictors of Immediate and Cumulative Benefit. Mindfulness 2021 Jan 15;12:2498–2509. [doi: https://doi.org/10.1007/s12671-021-01719-0]

13. Wetherill RR, Rao H, Hager N, Wang J, Franklin TR, Fan Y. Classifying and Characterizing Nicotine Use Disorder with High Accuracy Using Machine Learning and Resting-State fMRI. Addict Biol 2019 Jul;24(4):811–821. PMID:29949234

14. Watkins ER, Nolen-Hoeksema S. A habit-goal framework of depressive rumination. J Abnorm Psychol 2014 Feb;123(1):24–34. PMID:24661156

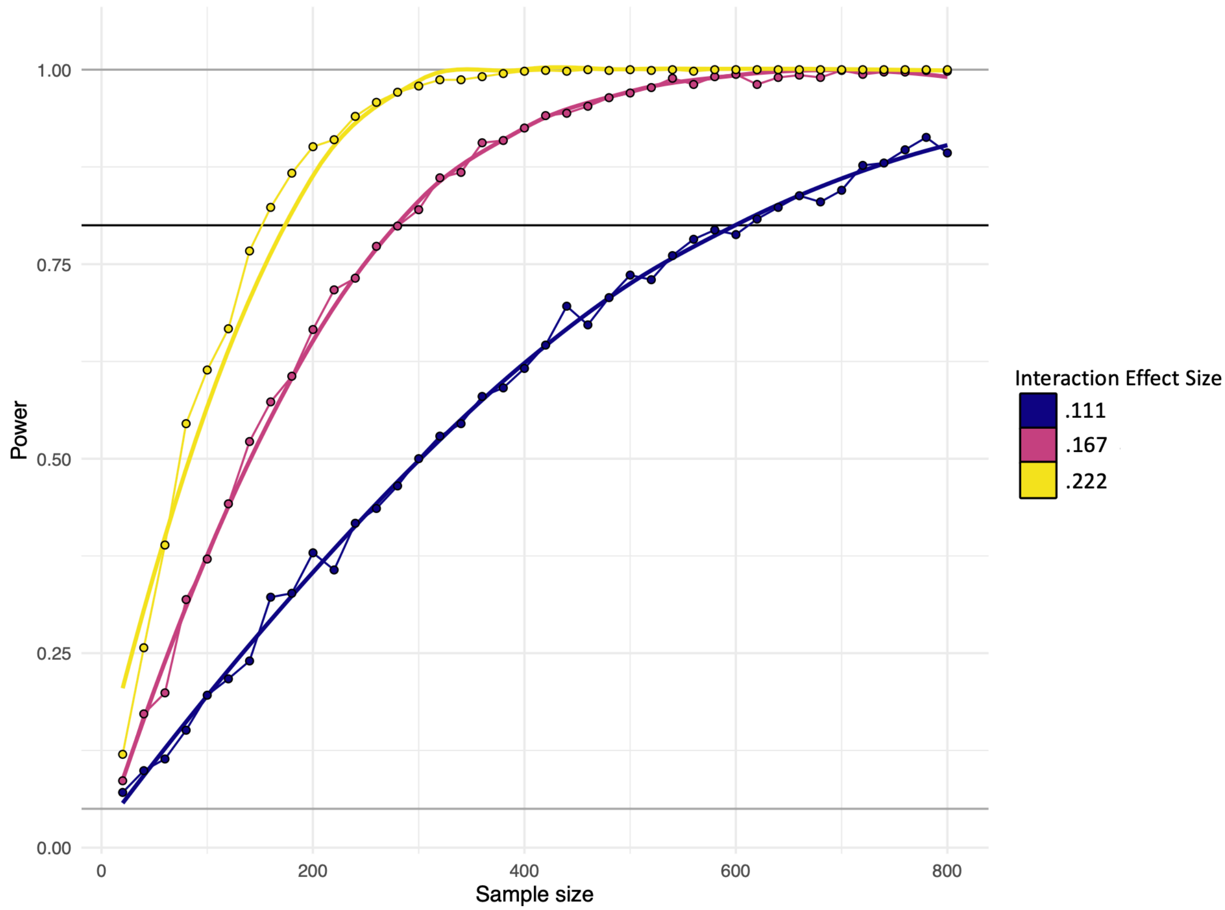
15. Watkins ER, Roberts H. Reflecting on rumination: Consequences, causes, mechanisms and treatment of rumination. Behav Res Ther 2020 Apr 1;127:103573. [doi: 10.1016/j.brat.2020.103573]

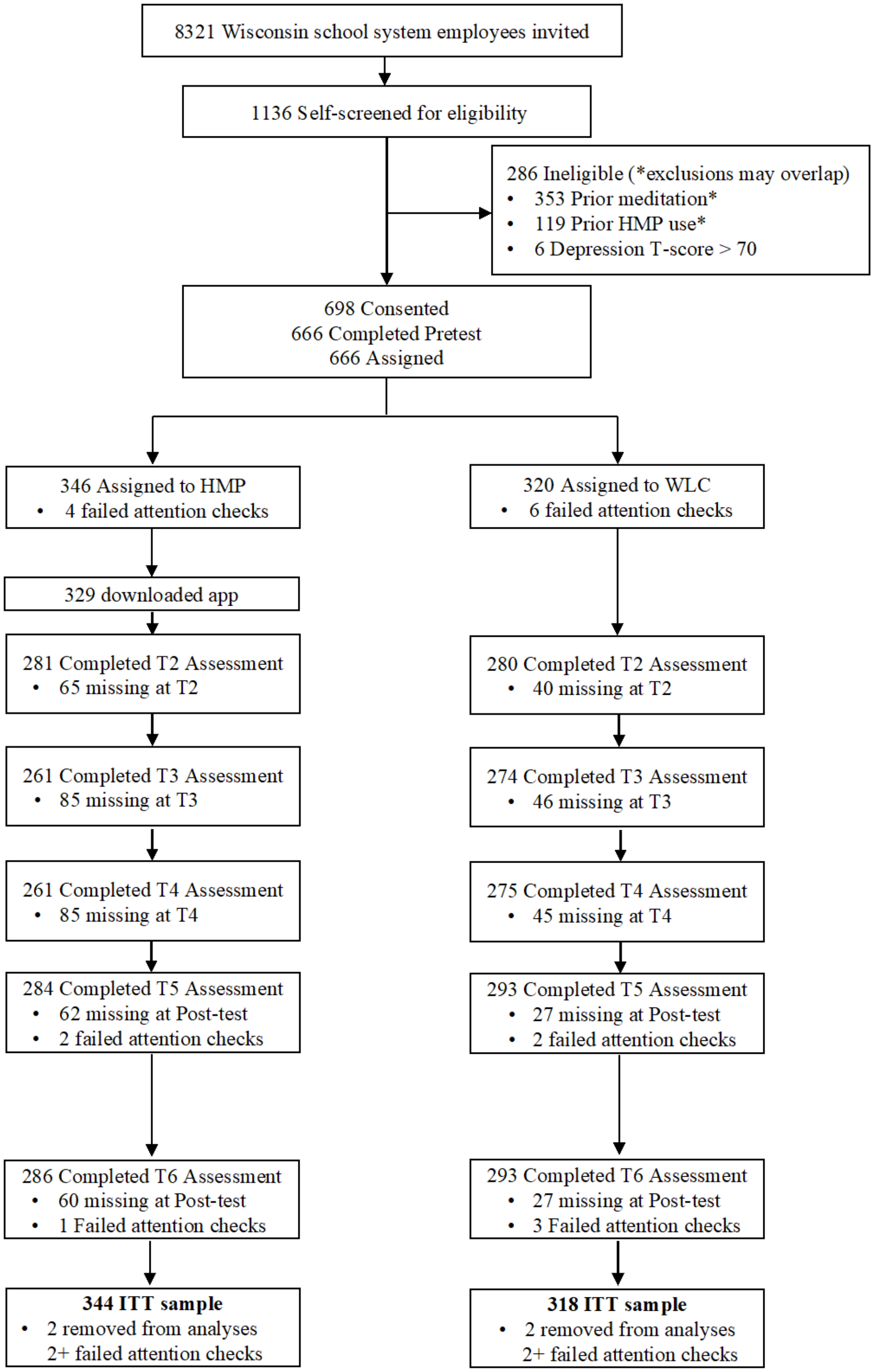
16. Webb CA, Swords CM, Lawrence H, Hilt LM. Which adolescents are well-suited to app-based mindfulness training? A randomized clinical trial and data-driven approach for personalized recommendations [Internet]. J Consult Clin Psychol. 2022 [cited 2021 Jul 21]. [doi: 10.31234/osf.io/7gbjd]

Supplemental Figures

Figure S1. Results of Monte Carlo simulations (1,000 iterations; implement via InteractionPowerR package in R). We inputted four *r*-type effect sizes (based on [16]) representing the correlation between the interaction term (Group x PAI) and outcome ($r = .22$), the correlation between Group and PAI scores ($r = - .12$) and their respective main effects ($r = -.21$; $r = .15$). The yellow curve plots power at varying sample sizes (80% power at n = 153). Two additional sets of simulations were conducted assuming an interaction effect size 75% (magenta; 80% power at n = 278) and 50% (blue; 80% power at n = 599) the magnitude of the above effect size. The black horizontal line represents 80% power.

Figure S2. CONSORT Flow Diagram. HMP = Healthy Minds Program; WLC = waitlist control (i.e., assessment only control); T1 = baseline, T2 = week 1, T3 = week 2, T4 = week 3, T5 = post-test, T6 = 3-month follow-up. Measures administered at these timepoints included clinical scales (depression, anxiety, stress, and loneliness), as well as measures of mindfulness, cognitive defusion, and meaning in life. For additional clinical trial and participant details (including CONSORT Checklist) see [4], https://psyarxiv.com/hrvmu/


1.00
0.75
0.50
0.25
0.00
Power
0
200
400
600
800
Sample size
Interaction Effect Size
.111
.167
.222


8321 Wisconsin school system employees invited
1136 Self-screened for eligibility
286 Ineligible (*exclusions may overlap)
• 353 Prior meditation*
• 119 Prior HMP use*
• 6 Depression T-score > 70
698 Consented
666 Completed Pretest
666 Assigned
346 Assigned to HMP
• 4 failed attention checks
320 Assigned to WLC
• 6 failed attention checks
329 downloaded app
281 Completed T2 Assessment
• 65 missing at T2
280 Completed T2 Assessment
• 40 missing at T2
261 Completed T3 Assessment
• 85 missing at T3
274 Completed T3 Assessment
• 46 missing at T3
261 Completed T4 Assessment
• 85 missing at T4
275 Completed T4 Assessment
• 45 missing at T4
284 Completed T5 Assessment
• 62 missing at Post-test
• 2 failed attention checks
293 Completed T5 Assessment
• 27 missing at Post-test
• 2 failed attention checks
286 Completed T6 Assessment
• 60 missing at Post-test
• 1 Failed attention checks
293 Completed T6 Assessment
• 27 missing at Post-test
• 3 Failed attention checks
344 ITT sample
• 2 removed from analyses
2+ failed attention checks
318 ITT sample
• 2 removed from analyses
2+ failed attention checks