# Personalized prediction of response to smartphone-delivered meditation training

Christian A. Webb[1], Matthew J. Hirshberg[2], Richard J. Davidson[2,3,4], Simon B. Goldberg[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

*Draft version 1, 7/29/21. This paper is currently under peer review. Please do not copy or cite without author's permission.*

**Keywords:** precision medicine; prediction; machine learning; meditation; mobile technology; personalized advantage index

**Abstract**

Meditation apps are popular and may reduce psychological distress, including during the COVID-19 pandemic. However, it is not clear who is most likely to benefit. Using randomized controlled trial data comparing a 4-week meditation app (Healthy Minds Program; HMP) with an assessment-only control 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 (preregistered primary outcome) from HMP vs. control. Significant Group x PAI interactions emerged, indicating that PAI scores moderated group differences in outcome. A regression model including repetitive negative thinking as the sole predictor performed comparably well. Finally, we demonstrate the translation of predictive models to personalized recommendations of expected benefit, which could inform users' decisions of whether to engage with a meditation app.

Precision medicine, which involves the use of individual variability to guide prevention and treatment (Collins & Varmus, 2015), has gained momentum within the health sciences in the past several years. 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 (e.g., in non-small-cell lung cancer, see Paez et al., 2004; Rosell et al., 2012; Schwaederle et al., 2015).

Psychiatry and clinical psychology have long hoped to better match patients with interventions. Over 50 years ago, Paul (1967) expressed the importance of determining "what treatment, by whom, is most effective for this individual with that specific problem, and under which circumstances?" (p. 111). Numerous studies have since examined patient-level factors as predictors of treatment response (Kessler et al., 2017; Norcross & Wampold, 2018). Candidate characteristics associated with either overall response (i.e., prognostic predictors) or differential response (i.e., prescriptive predictors; Cohen & DeRubeis, 2018) include sociodemographic factors (e.g., age, employment, education), symptom severity, psychiatric comorbidities, personality factors, attachment style, and history of adversity (Kessler et al., 2017). 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 (Hastie et al., 2017; Webb & Cohen, 2021). 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 (Boehmke & Greenwell, 2019). 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; Dwyer et al., 2018). 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 (Aafjes-van Doorn et al., 2021; Chekroud et al., 2021; Dwyer et al., 2018). Previous applications have included the prediction of response to psychiatric medications (e.g., Chekroud et al., 2016, 2017; Hayes et al., 2016), evidence-based psychotherapies (Chien et al., 2020; Deisenhofer et al., 2018; Delgadillo & Gonzalez Salas Duhne, 2020; Webb et al., 2020), and other non-pharmacological approaches such as transcranial magnetic stimulation (Corlier et al., 2019).

In pursuit of precision mental health (Cohen & DeRubeis, 2018; Delgadillo & Lutz, 2020; DeRubeis, 2019), researchers have leveraged machine learning approaches in an effort to optimize treatment recommendations. For example, DeRubeis et al. (2014) 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 (DeRubeis et al., 2014), CBT vs. interpersonal therapy (Huibers et al., 2015), CBT vs.

psychodynamic therapy (Cohen et al., 2019), antidepressant medication vs. placebo (Webb et al., 2018), and supportive expressive therapy vs. an antidepressant vs. placebo (Zilcha-Mano et al., 2016).

Prior research using the PAI and related approaches (Chekroud et al., 2021) 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 (Jorm et al., 2017; Kohn et al., 2004). Digital health technologies, such as internet-based CBT (Andersson & Cuijpers, 2009) and smartphone-delivered mental health apps (Linardon et al., 2019), have the potential to substantially increase access to evidence-based treatments (Steinhubl et al., 2013). 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 (Torous et al., 2019). 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 (Chien et al., 2020; Eysenbach, 2005). Moreover, the scalability of mHealth makes collection of adequately powered sample sizes for robust modeling a possibility (Luedtke et al., 2019).

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 (Lagan et al., 2021). 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 (Wasil et al., 2020). Despite the soaring popularity of meditation apps, a critical question remains unanswered: *For whom* is app-based mindfulness 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 (Hirshberg et al., 2021). The RCT was conducted in a sample of school district employees ($n$ = 662) in the state of Wisconsin during the COVID-19 pandemic. Using data from this RCT, the overarching goal of this study was to develop and evaluate a data driven (PAI) approach to inform personalized meditation app recommendations. 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 Hirshberg et al., 2021). Eligible participants were adults (≥18 years of age) currently employed by a Wisconsin school who owned a smartphone capable of downloading the HMP, who 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; Pilkonis et al., 2011). A total of 666 participants were enrolled, and 4 were removed for failing multiple attention checks, resulting in a sample size of 662 for these analyses. Upon completing pre-test measures, participants were randomly assigned (1:1) through the Research Electronic Data Capture (REDCap) system to use the four-week HMP or an assessment-only control condition. 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). Participants were paid US$150 for completing all assessments.

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 R code 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 (Dahl et al., 2020; Goldberg et al., 2020). 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. For additional details on HMP app content, see Hirshberg et al. (2021).

## 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 (Pilkonis et al., 2011) and the 10-item Perceived Stress Scale (PSS; Cohen, 1988). All three are widely used measures with established reliability and validity (Roberti et al., 2006; Schalet et al., 2016). Internal consistency of the PSS was adequate in the current sample ($\alpha$ = .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]). 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; Ehring et al., 2011) assessed worry and rumination ($\alpha$ = .95). The five-item World Health Organization (WHO)-5 (Topp et al., 2015) assessed global well-being ($\alpha$ = .85). The eight-item Act with Awareness subscale of the Five Facet Mindfulness Questionnaire ($\alpha$ = .91, FFMQ; Baer et al., 2008) assessed mindful attention in daily life. The five-item National Institutes of Health (NIH) Toolbox Loneliness Questionnaire (NIHTL; Cyranowski et al., 2013) assessed perceived social disconnection ($\alpha$ = .90). The 12-item Self-Compassion Scale Short Form (SCSSF; Raes et al., 2011) assessed feelings of kindness towards oneself ($\alpha$ = .86). The 10-item Drexel Defusion Scale (DDS; Forman et al., 2012) assessed ability to experientially distance from internal experiences ($\alpha$ = .84). The 10-item Meaning in Life Questionnaire (MLQ; Steger et al., 2006) assessed presence and search for meaning ($\alpha$s = .91 and .93, respectively). Baseline scores on symptom measures (PROMIS Depression, PROMIS Anxiety, PSS, and the composite distress scale) and several demographic variables (age, gender, race, marital status, and income) were also included as predictors.

**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; Stekhoven & Bühlmann, 2012). 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%). Prior to 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 (Steyerberg, 2019).

***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; Moons et al., 2015; Varoquaux et al., 2017; Webb, Swords, et al., 2021; Wetherill et al., 2019). 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, Friedman et al., 2010) 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 (Kuhn, 2008) 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.

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 (e.g., Cohen et al., 2019; Schwartz et al., 2020; Webb et al., 2018), 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). One approach to evaluating PAI predictions is to statistically compare (e.g., via two-tailed t-tests) observed outcomes for those participants randomly assigned to their PAI-indicated condition (i.e., HMP for PAI scores $< 0$ and control condition for PAI scores $> 0$) vs. the contraindicated condition (Cohen et al., 2019; DeRubeis et al., 2014; Huibers et al., 2015; Schwartz et al., 2020; Webb et al., 2018). However, given that PAI scores revealed that nearly all participants were predicted to have better outcomes in HMP relative to control (see **Meditation App Recommendations** below) this approach was not used. In addition, the latter approach dichotomizes an inherently continuous variable (PAI). Instead, we tested whether PAI scores (as a continuous variable) moderated treatment group differences in outcome (i.e., slope of change in distress) via a Group (i.e., intervention condition) x PAI interaction (while adjusting for baseline distress scores). 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, 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 (Webb, Swords, et al., 2021a, 2021b). 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; R Core Team, 2021). 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 (Webb, Swords, et al., 2021b) 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 = 0.5 and power = 80%; see Supplemental Figure S1, including Figure note, for additional power analysis details).

## Results

### Sample Demographics

Within the full sample ($n$ = 662), 344 were assigned to the HMP condition and 318 to the assessment-only control group. 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$150,000 (30.4%). The majority (79.0%) reported depression and/or anxiety symptoms at baseline that were above the clinical cutoff on the PROMIS Depression and PROMIS Anxiety measures ($T > 55$).

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**). 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 (adjusting for baseline distress scores) in predicting outcome ($t$(657) = 3.07, $p$ = .003; adjusted $r^2$ = .14), 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 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 (adjusting for baseline distress scores) in predicting symptom change ($t(657) = 3.65$, $p < .001$; adjusted $r^2 = .15$), indicating that PAI scores moderated group differences in outcome (**Figure 2**). Specifically, as PAI scores decreased (reflecting increasing PTQ repetitive negative thinking scores) group differences favoring the HMP condition also increased. Given the association between repetitive negative thinking and depressive symptoms (Watkins & Nolen-Hoeksema, 2014; Watkins & Roberts, 2020), we also conducted additional sensitivity analyses controlling for baseline levels of depressive symptoms, 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 near equivalent performance relative to a more complex multivariable ENR model (i.e., adjusted $r^2 = .15$ vs. $r^2 = .14$, 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; Canty & Ripley, 2021). 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 (Hayes & Matthes, 2009)(Interactions package; Long, 2019) 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 **Figures 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 50$^{th}$ percentile, they would be predicted to be at the 41$^{st}$ percentile (relative to pre-intervention distress scores) following the 4-week mindfulness app course vs. 58$^{th}$ 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 (Wasil et al., 2020). Despite their growing popularity, little is known regarding who in fact benefits from these apps. In the current study, we developed an algorithm (i.e., PAI) 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.

Consistent with prior studies in psychotherapy (e.g., DeRubeis et al., 2014; Huibers et al., 2015) and pharmacotherapy (e.g., Webb et al., 2018), we found evidence that a machine learning-based PAI 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 (Hirshberg et al., 2021), 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" in **Figure 3**). 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.

Importantly, 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. Repetitive negative thinking moderated outcome to app-based meditation training relative to the assessment-only control. Similar to two prior studies in adolescents with elevated rumination (Webb, Swords, et al., 2021a, 2021b), 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, such as different versions of a single app (e.g., Goldberg et al., 2020; Roepke et al., 2015). 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, 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; Flett et al., 2019; Lagan et al., 2021), or even compare mHealth interventions vs.

conventional (in-person) psychotherapy or pharmacotherapy. Such studies could determine, for example, whether we can predict which individuals with depressive symptoms require more intensive, weekly CBT (or a longer-term antidepressant prescription) vs. those who would experience symptom remission from a brief meditation or CBT-based app course.

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; Lipton et al., 2000). In theory, this latter approach could minimize delay to receiving an effective intervention.

As perhaps with any implementation of machine learning within health care, this possibility – using data-driven algorithms for treatment matching – raises important ethical considerations. Chekroud et al. (2021) highlight several key ethical challenges facing the deployment of machine learning in clinical settings. One particularly germane to the current study is the possibility that algorithms may not work equally well for all individuals. Models can be expected to perform worse for demographic groups underrepresented in training data (in our example, racial/ethnic minorities and males). An inadequately trained model could provide predictions for underrepresented groups that are at best inaccurate and at worst systematically biased towards recommending suboptimal treatment approaches (Obermeyer et al., 2019). Thus, it is vitally important that researchers and clinicians involved in these efforts remain acutely sensitive to the possibility of replicating and/or accentuating preexisting systematic inequities.

**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; Chekroud et al., 2021). 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., Webb et al., 2021; Webb, Israel, et al., 2021). Second, we were unable to conduct external validation by evaluating performance in an entirely new sample (e.g., from another RCT). Third, we did not include an active comparison condition. 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 PAI 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 PAI-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 Constantino et al., 2021). 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 (Kessler et al., 2017). 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 (Bone et al., 2021; Delgadillo et al., 2020; Webb et al., 2020; Webb, Forgeard, et al., 2021). 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 (Luedtke et al., 2019). 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.

**References**

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2021). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research*, *31*(1), 92–116.

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

Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., Walsh, E., Duggan, D., & Williams, J. M. G. (2008). Construct Validity of the Five Facet Mindfulness Questionnaire in Meditating and Nonmeditating Samples. *Assessment*, *15*(3), 329–342. https://doi.org/10.1177/1073191107313003

Boehmke, B., & Greenwell, B. M. (2019). *Hands-On Machine Learning with R* (1st edition). Chapman and Hall/CRC.

Bone, C., Simmonds-Buckley, M., Thwaites, R., Sandford, D., Merzhvynska, M., Rubel, J., Deisenhofer, A.-K., Lutz, W., & Delgadillo, J. (2021). Dynamic prediction of psychological treatment outcomes: Development and validation of a prediction model using routinely collected symptom data. *The Lancet Digital Health*, *3*(4), e231–e240.

Canty, A., & Ripley, B. (2021). *boot: Bootstrap R (S-Plus) Functions.*

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., & Iniesta, R. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154–170.

Chekroud, A. M., Gueorguieva, R., Krumholz, H. M., Trivedi, M. H., Krystal, J. H., & McCarthy, G. (2017). Reevaluating the efficacy and predictability of antidepressant treatments: A symptom clustering approach. *JAMA Psychiatry*, *74*(4), 370–378.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *The Lancet Psychiatry*, *3*(3), 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X

Chien, I., Enrique, A., Palacios, J., Regan, T., Keegan, D., Carter, D., Tschiatschek, S., Nori, A., Thieme, A., & Richards, D. (2020). A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. *JAMA Network Open*, *3*(7), e2010791–e2010791.

Cohen, S. (1988). *Perceived stress in a probability sample of the United States.*

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment Selection in Depression. *Annual Review of Clinical Psychology*, *14*(1), 209–236. https://doi.org/10.1146/annurev-clinpsy-050817-084746

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2019). 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*, *0*(0), 1–14. https://doi.org/10.1080/10503307.2018.1563312

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, *372*(9), 793–795.

Constantino, M. J., Boswell, J. F., Coyne, A. E., Swales, T. P., & Kraus, D. R. (2021). Effect of Matching Therapists to Patients vs Assignment as Usual on Adult Psychotherapy Outcomes: A Randomized Clinical Trial. *JAMA Psychiatry*.

Corlier, J., Carpenter, L. L., Wilson, A. C., Tirrell, E., Gobin, A. P., Kavanaugh, B., & Leuchter, A. F. (2019). The relationship between individual alpha peak frequency and clinical outcome with repetitive Transcranial Magnetic Stimulation (rTMS) treatment of Major Depressive Disorder (MDD). *Brain Stimulation*, *12*(6), 1572–1578.

Cyranowski, J. M., Zill, N., Bode, R., Butt, Z., Kelly, M. A., Pilkonis, P. A., Salsman, J. M., & Cella, D. (2013). Assessing social support, companionship, and distress: National Institute of Health (NIH) Toolbox Adult Social Relationship Scales. *Health Psychology*, *32*(3), 293.

Dahl, C. J., Wilson-Mendenhall, C. D., & Davidson, R. J. (2020). The plasticity of well-being: A training-based framework for the cultivation of human flourishing. *Proceedings of the National Academy of Sciences*, *117*(51), 32197–32206.

Deisenhofer, A.-K., Delgadillo, J., Rubel, J. A., Boehnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety*, *35*(6), 541–550.

Delgadillo, J., Appleby, S., Booth, S., Burnett, G., Carey, A., Edmeade, L., Green, S., Griffin, P., Johnson, E., & Jones, R. (2020). The Leeds Risk Index: Field-test of a stratified psychological treatment selection algorithm. *Psychotherapy and Psychosomatics*.

Delgadillo, J., & Gonzalez Salas Duhne, P. (2020). Targeted prescription of cognitive–behavioral therapy versus person-centered counseling for depression using a machine learning approach. *Journal of Consulting and Clinical Psychology*, *88*(1), 14.

Delgadillo, J., & Lutz, W. (2020). A development pathway towards precision mental health care. *JAMA Psychiatry*, *77*(9), 889–890.

DeRubeis, R. J. (2019). The history, current status, and possible future of precision mental health. *Behaviour Research and Therapy*, *123*, 103506. https://doi.org/10.1016/j.brat.2019.103506

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment Recommendations. A Demonstration. *PLOS ONE*, *9*(1), e83875. https://doi.org/10.1371/journal.pone.0083875

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118.

Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., & Ehlers, A. (2011). The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(2), 225–232.

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research*, *7*(1), e11.

Flett, J. A., Hayne, H., Riordan, B. C., Thompson, L. M., & Conner, T. S. (2019). Mobile Mindfulness Meditation: A Randomised Controlled Trial of the Effect of Two Popular Apps on Mental Health. *Mindfulness*, *10*(5), 863–876.

Forman, E. M., Herbert, J. D., Juarascio, A. S., Yeomans, P. D., Zebell, J. A., Goetter, E. M., & Moitra, E. (2012). The Drexel defusion scale: A new measure of experiential distancing. *Journal of Contextual Behavioral Science*, *1*(1–2), 55–65.

Goldberg, S. B., Imhoff-Smith, T., Bolt, D. M., Wilson-Mendenhall, C. D., Dahl, C. J., Davidson, R. J., & Rosenkranz, M. A. (2020). Testing the Efficacy of a Multicomponent, Self-Guided, Smartphone-Based Meditation App: Three-Armed Randomized Controlled Trial. *JMIR Mental Health*, *7*(11), e23825. https://doi.org/10.2196/23825

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* (2nd edition). Springer.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, *41*(3), 924–936.

Hayes, J. F., Marston, L., Walters, K., Geddes, J. R., King, M., & Osborn, D. P. (2016). Lithium vs. Valproate vs. Olanzapine vs. Quetiapine as maintenance monotherapy for bipolar disorder: A population-based UK cohort study using electronic health records. *World Psychiatry*, *15*(1), 53–58.

Hirshberg, M. J., Frye, C., Dahl, C. J., Riordan, K., Vack, N., Sachs, J., Goldman, R. I., Davidson, R. J., & Goldberg, S. (2021). *A Randomized Controlled Trial of a Smartphone-Based Well-Being Training in Public School System Employees During the COVID-19 Pandemic*. PsyArXiv. https://doi.org/10.31234/osf.io/hrvmu

Huibers, M. J. H., Cohen, Z. D., Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting Optimal Outcomes in Cognitive Therapy or Interpersonal Psychotherapy for Depressed Individuals Using the Personalized Advantage Index Approach. *PLOS ONE*, *10*(11), e0140771. https://doi.org/10.1371/journal.pone.0140771

Jorm, A. F., Patten, S. B., Brugha, T. S., & Mojtabai, R. (2017). Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry*, *16*(1), 90–99.

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., de Jonge, P., Nierenberg, A. A., Rosellini, A. J., Sampson, N. A., Schoevers, R. A., Wilcox, M. A., & Zaslavsky, A. M. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, *26*(1), 22–36. https://doi.org/10.1017/S2045796016000020

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

Lagan, S., D'Mello, R., Vaidyam, A., Bilden, R., & Torous, J. (2021). Assessing Mental Health Apps Marketplaces with Objective Metrics from 29,190 Data Points from 278 Apps. *Acta Psychiatrica Scandinavica*.

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019). The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. *World Psychiatry*, *18*(3), 325–336.

Lipton, R. B., Stewart, W. F., Stone, A. M., Láinez, M. J., & Sawyer, J. P. (2000). Stratified care vs step care strategies for migraine: The Disability in Strategies of Care (DISC) Study: A randomized trial. *Jama*, *284*(20), 2599–2605.

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample Size Requirements for Multivariate Models to Predict Between-Patient Differences in Best Treatments of Major Depressive Disorder. *Clinical Psychological Science*, *7*(3), 445–461. https://doi.org/10.1177/2167702618815466

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*, *162*(1), W1–W73.

Norcross, J. C., & Wampold, B. E. (2018). A new therapy for each patient: Evidence-based relationships and responsiveness. *Journal of Clinical Psychology*, *74*(11), 1889–1906.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453.

Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., & Boggon, T. J. (2004). EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science*, *304*(5676), 1497–1500.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*(2), 109.

Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., Cella, D., & Group, P. C. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): Depression, anxiety, and anger. *Assessment*, *18*(3), 263–283.

Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the self-compassion scale. *Clinical Psychology & Psychotherapy*, *18*(3), 250–255.

Roberti, J. W., Harrington, L. N., & Storch, E. A. (2006). Further psychometric support for the 10-item version of the perceived stress scale. *Journal of College Counseling*, *9*(2), 135–147.

Rosell, R., Carcereny, E., Gervais, R., Vergnenegre, A., Massuti, B., Felip, E., Palmero, R., Garcia-Gomez, R., Pallares, C., & Sanchez, J. M. (2012). 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*, *13*(3), 239–246.

Schalet, B. D., Pilkonis, P. A., Yu, L., Dodds, N., Johnston, K. L., Yount, S., Riley, W., & Cella, D. (2016). Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. *Journal of Clinical Epidemiology*, *73*, 119–127.

Schwaederle, M., Zhao, M., Lee, J. J., Eggermont, A. M., Schilsky, R. L., Mendelsohn, J., Lazar, V., & Kurzrock, R. (2015). Impact of precision medicine in diverse cancers: A meta-analysis of phase II clinical trials. *Journal of Clinical Oncology*, *33*(32), 3817.

Schwartz, B., Cohen, Z. D., Rubel, J. A., Zimmermann, D., Wittmann, W. W., & Lutz, W. (2020). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research*, 1–19.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology*, *53*(1), 80.

Steinhubl, S. R., Muse, E. D., & Topol, E. J. (2013). Can mobile health technologies transform health care? *Jama*, *310*(22), 2395–2396.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Steyerberg, E. W. (2019). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating* (2nd ed. 2019 edition). Springer.

Team. (2021). R: A language and environment for statistical computing. *Vienna, Austria: R Foundation for Statistical Computing*. https://www.R-project.org/

Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, *84*(3), 167–176.

Torous, J., Andersson, G., Bertagnoli, A., Christensen, H., Cuijpers, P., Firth, J., Haim, A., Hsin, H., Hollis, C., & Lewis, S. (2019). Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry*, *18*(1), 97.

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, *145*, 166–179. https://doi.org/10.1016/j.neuroimage.2016.10.038

Wasil, A. R., Gillespie, S., Shingleton, R., Wilks, C. R., & Weisz, J. R. (2020). Examining the Reach of Smartphone Apps for Depression and Anxiety. *American Journal of Psychiatry*, *177*(5), 464–465. https://doi.org/10.1176/appi.ajp.2019.19090905

Watkins, E. R., & Nolen-Hoeksema, S. (2014). A habit-goal framework of depressive rumination. *Journal of Abnormal Psychology*, *123*(1), 24–34. https://doi.org/10.1037/a0035540

Watkins, E. R., & Roberts, H. (2020). Reflecting on rumination: Consequences, causes, mechanisms and treatment of rumination. *Behaviour Research and Therapy*, *127*, 103573. https://doi.org/10.1016/j.brat.2020.103573

Webb, C. A., & Cohen, Z. D. (2021). Progress towards clinically informative data-driven decision support tools in psychotherapy. *The Lancet Digital Health*, *3*(4), e207–e208.

Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). 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*, *88*(1), 25–38. https://doi.org/10.1037/ccp0000451

Webb, C. A., Forgeard, M., Israel, E. S., Lovell-Smith, N., Beard, C., & Björgvinsson, T. (2021). Personalized prescriptions of therapeutic skills from patient characteristics: An ecological momentary assessment approach. *Journal of Consulting and Clinical Psychology*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/ccp0000555

Webb, C. A., Israel, E. S., Belleau, E., Appleman, L., Forbes, E. E., & Pizzagalli, D. A. (2021). 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*, *60*(3), 377–387. https://doi.org/10.1016/j.jaac.2020.03.010

Webb, C. A., Swords, C. M., Lawrence, H., & Hilt, L. M. (2021b). *Which adolescents are well-suited to app-based mindfulness training? A randomized clinical trial and data-driven approach for personalized recommendations*. PsyArXiv. https://doi.org/10.31234/osf.io/7gbjd

Webb, C. A., Swords, C. M., Murray, L. M., & Hilt, L. M. (2021a). *App-based Mindfulness Training for Adolescent Rumination: Predictors of Immediate and Cumulative Benefit*. PsyArXiv. https://doi.org/10.31234/osf.io/k4ygw

Webb, C. A., Trivedi, M. H., Cohen, Z. D., Dillon, D. G., Fournier, J. C., Goer, F., Fava, M., McGrath, P. J., Weissman, M., Parsey, R., Adams, P., Trombello, J. M., Cooper, C., Deldin, P., Oquendo, M. A., McInnis, M. G., Huys, Q., Bruder, G., Kurian, B. T., … Pizzagalli, D. A. (2018). Personalized prediction of antidepressant v. placebo response: Evidence from the EMBARC study. *Psychological Medicine*, 1–10. https://doi.org/10.1017/S0033291718001708

Wetherill, R. R., Rao, H., Hager, N., Wang, J., Franklin, T. R., & Fan, Y. (2019). Classifying and Characterizing Nicotine Use Disorder with High Accuracy Using Machine Learning and Resting-State fMRI. *Addiction Biology*, *24*(4), 811–821. https://doi.org/10.1111/adb.12644

Zilcha-Mano, S., Keefe, J., Chiu, H., Rubin, A., Barrett, M. S., & Barber, J. P. (2016). Reducing Dropout in Treatment for Depression: Translating Dropout Predictors Into Individualized Treatment Recommendations. *The Journal of Clinical Psychiatry*, *77*(12), e1584–e1590. https://doi.org/10.4088/JCP.15m10081

## Compliance with Ethical Standards

This study was approved by the University of Wisconsin-Madison IRB and was therefore performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

## Informed consent:

Informed consent was obtained from all individual participants included in the study.

## Author Contributions

CW and SG developed the study concept. MH, RD, and SG designed and conducted the parent trial. CW conceptualized, performed, and interpreted the data analysis with consultation from SG. CW and SG drafted the paper and MH and RD provided critical revisions. All authors approved the final version of the paper for submission.

## Author Note

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. RJD is the founder, president, and serves on the board of directors for the nonprofit organization, Healthy Minds Innovations, Inc.

**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).

**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**

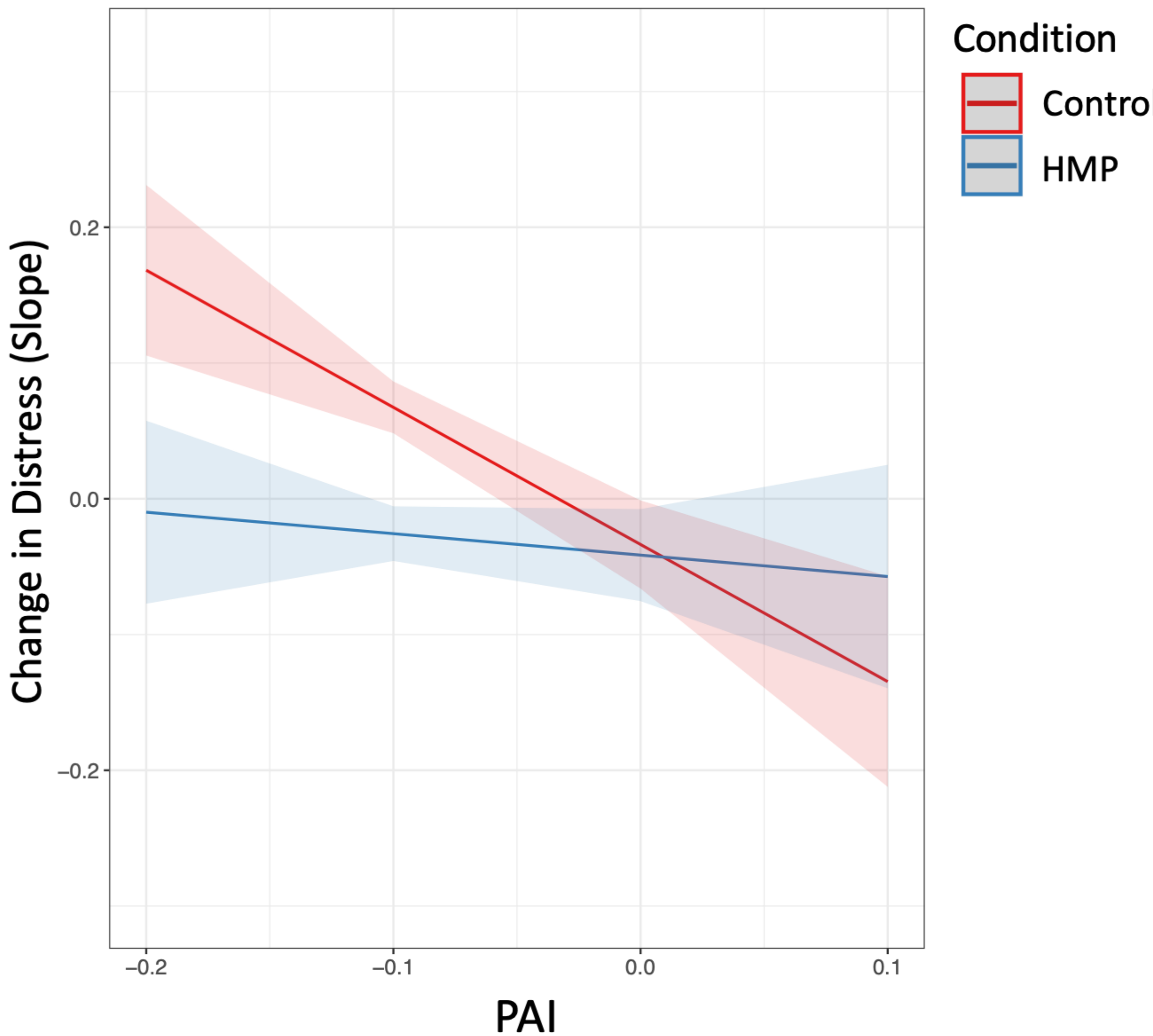

Condition
Control
HMP
Change in Distress (Slope)
0.2
0.0
−0.2
−0.2
−0.1
0.0
0.1
PAI

**Figure 2**

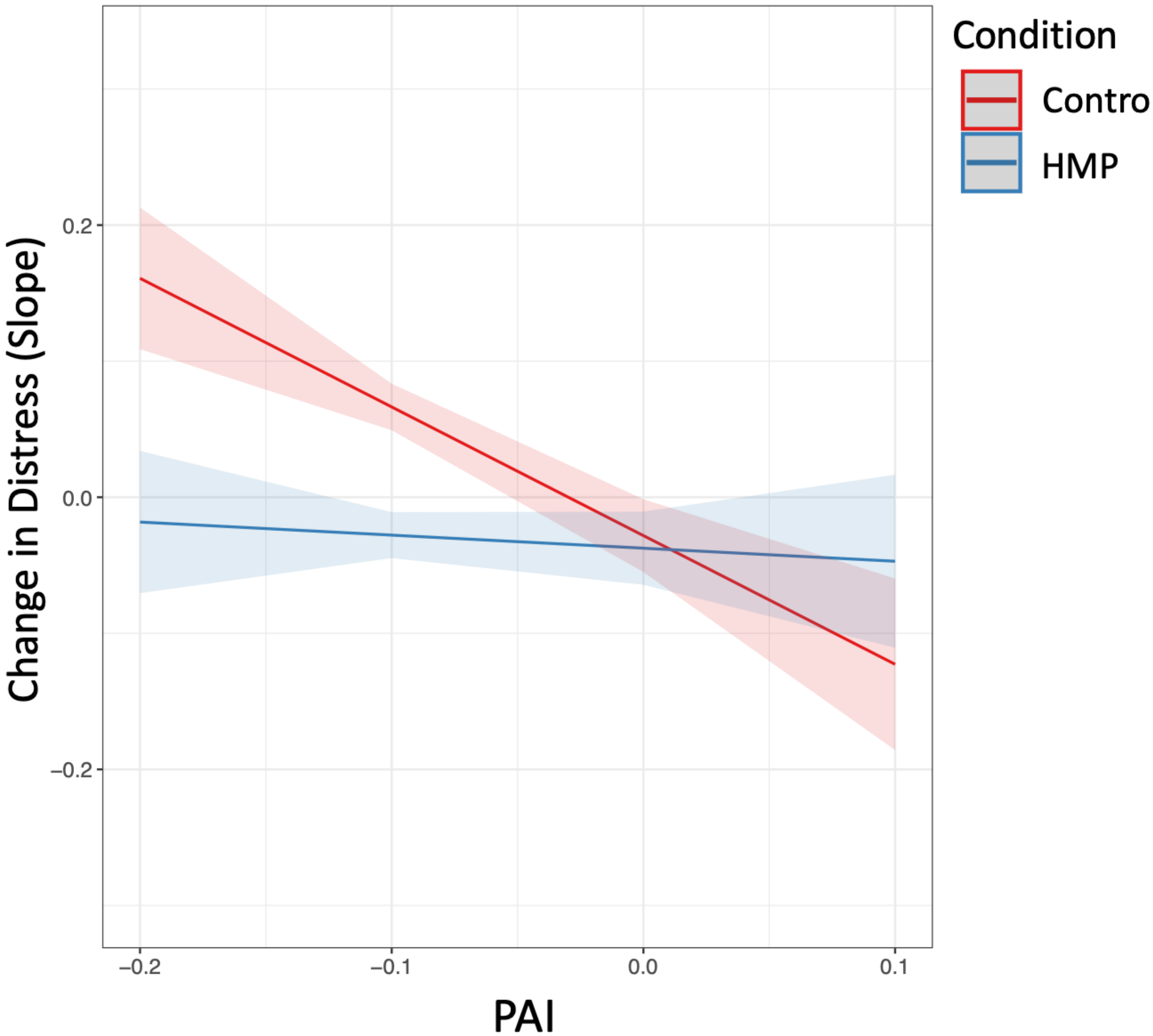
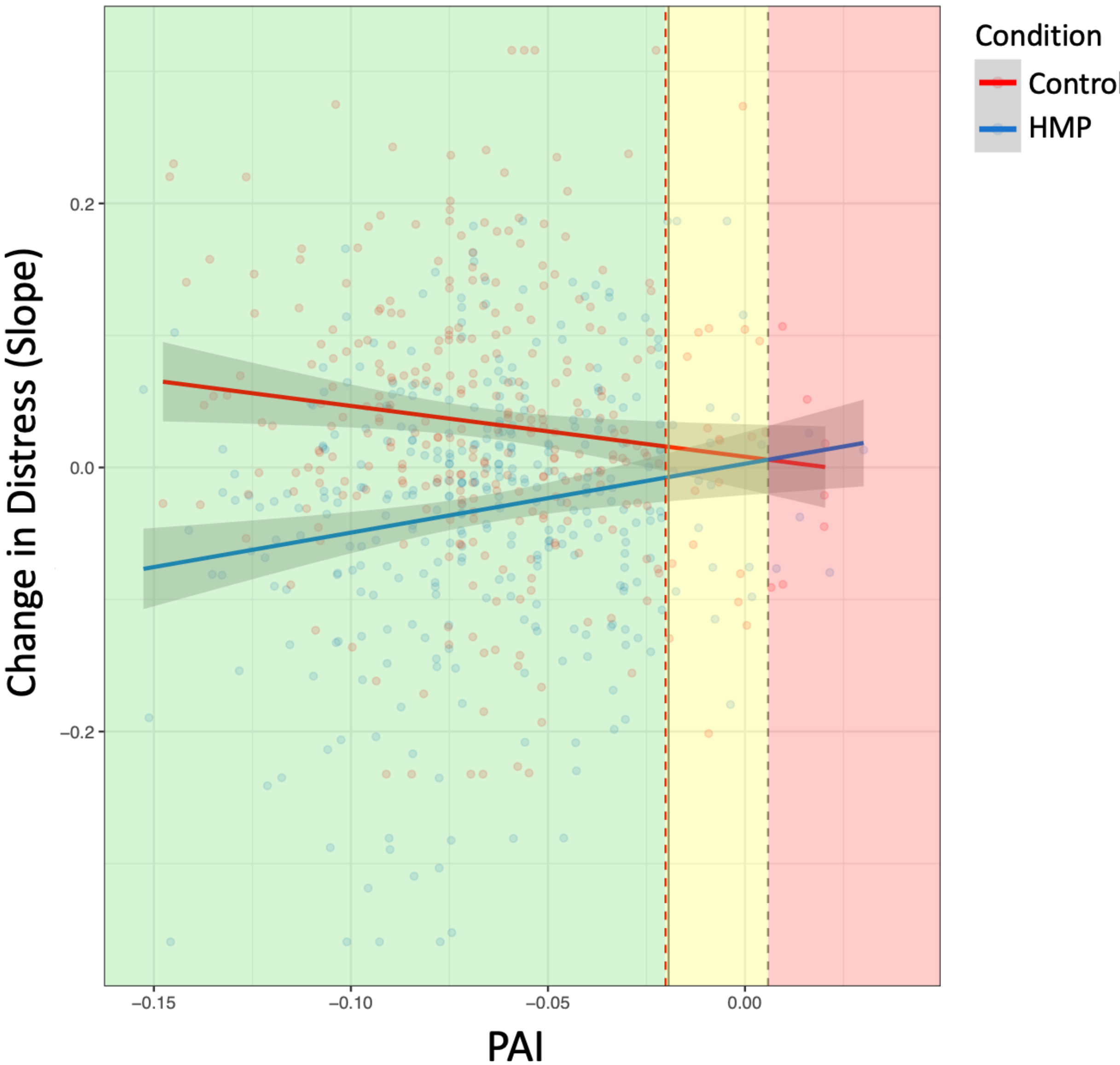

Condition
Control
HMP
Change in Distress (Slope)
0.2
0.0
−0.2
−0.2
−0.1
0.0
0.1
PAI

**Figure 3**


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