Behaviour Research and Therapy 205 (2026) 105141

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Personalized prediction of response to three meditation practices: A randomized controlled trial

Otto Simonsson[a,b,c,*], Christian A. Webb[d,e], Mini Ruiz[f], Walter Osika[a,c], Jayanth Narayanan[g], Micael Dahlén[c], Wolfgang Lutz[h], Brian Schwartz[h], Boyu Ren[d,e], Cortland J. Dahl[b,i], Richard J. Davidson[b,i,j,k], Matthew J. Hirshberg[b], Simon B. Goldberg[b,l,**]

[a] *Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden*
[b] *Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, USA*
[c] *Center for Wellbeing, Welfare and Happiness, Stockholm School of Economics, Stockholm, Sweden*
[d] *Harvard Medical School, Boston, MA, USA*
[e] *Center for Depression, Anxiety, and Stress Research, McLean Hospital, Belmont, MA, USA*
[f] *Clinical Sciences, Intervention and Technology (CLINTEC), Karolinska Institute, Stockholm, Sweden*
[g] *Department of Management & Organizational Development, Northeastern University, Boston, MA, USA*
[h] *Department of Psychology, Trier University, Trier, Germany*
[i] *Healthy Minds Innovations Inc, Madison, WI, USA*
[j] *Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA*
[k] *Department of Psychiatry, University of Wisconsin-Madison, Madison, WI, USA*
[l] *Department of Counseling Psychology, University of Wisconsin - Madison, Madison, WI, USA*

A R T I C L E I N F O

Keywords:
Meditation
Mindfulness
Gratitude
Self-compassion
Wellbeing
Machine learning
Personalized

A B S T R A C T

Despite decades of meditation research, there is a lack of preregistered and well-powered studies that investigate the effects of brief meditation practices on subjective wellbeing and use data-driven approaches to predict response to these interventions. Using representative samples of the US and UK adult populations with regards to age, sex, and ethnicity ($N = 5049$), we conducted a preregistered and well-powered randomized controlled trial to investigate the effects of three brief (5-10 min) meditation practices (i.e., mindfulness, self-compassion, gratitude) relative to a control condition. We also examined whether a data-driven approach can predict an individual's response to these interventions from their pre-intervention characteristics. Taken together, our findings suggest that brief meditation practices modestly improve an aspect of subjective wellbeing (negative affect) and such response can be predicted using pre-intervention characteristics. Future prospective randomized trials should examine algorithm-guided intervention assignment to directly test the potential of personalized meditation recommendations to improve subjective wellbeing.

1. Introduction

There has been growing scientific interest in meditation-based interventions designed to cultivate specific qualities of mind (e.g., mindfulness, self-compassion, gratitude) as a means to improve subjective wellbeing (Folk & Dunn, 2023; Goldberg & Davidson, 2024; van Agteren et al., 2021), which typically comprises positive affect, negative affect, and life satisfaction (Deiner, 1984; Deiner et al., 1999). However, despite extensive research on the effects of meditation-based interventions on subjective wellbeing measures (e.g., positive and negative affect; Folk & Dunn, 2023), these studies have often lacked methodological rigor and have rarely been *both* preregistered and well-powered (Folk & Dunn, 2023). This increases the risk of Type I errors due to lack of preregistration and Type II errors due to insufficient power, which underscores the importance of conducting preregistered and well-powered studies to ensure reliable and reproducible findings (Ioannidis, 2005; Maxwell, 2004; Nosek et al., 2018).

Notably, a recent preregistered and well-powered randomized controlled trial of a digital meditation-based wellbeing training ($n = 662$) found significant intervention effects on subjective wellbeing,

* Corresponding author. Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden.
** Corresponding author. Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, USA.
E-mail addresses: otto.simonsson@ki.se (O. Simonsson), sbgoldberg@wisc.edu (S.B. Goldberg).

https://doi.org/10.1016/j.brat.2026.105141
Received 10 June 2025; Received in revised form 10 July 2026; Accepted 9 August 2026
Available online 10 August 2026

but because multiple meditation practices were included in the digital intervention, it is not clear whether certain meditation practices contributed to a greater extent to the observed improvements in subjective wellbeing (Hirshberg et al., 2022). By testing individual practices as separate experimental conditions, the effects of each can be estimated. Such granular research could, for example, clarify which individual components within meditation-based interventions are most effective, informing the design of more precise and effective interventions. While attempts to apply this approach in meditation research remain limited (e.g., Lindsay et al., 2018), precedent exists in the broader psychotherapy literature, where dismantling studies have informed the development of more refined treatments (e.g., the dismantling of cognitive behavioral therapy that contributed to behavioral activation; Jacobson et al., 1996; see also Bell et al., 2013).

Although a meta-analysis of 54 effect sizes from randomized trials testing brief meditation-based interventions showed decreases in negative affect relative to controls (Hedges' $g = 0.22$), it also showed evidence for publication bias (i.e., not publishing null findings; Schumer et al., 2018), which increases the risk of inflated estimates of intervention effects. Other types of bias common in meta-analyses (e.g., selective outcome reporting bias) may similarly contribute to inflated meta-analytic effect sizes (Kvarven et al., 2020). Because of these biases in meta-analyses, it is important to conduct preregistered and well-powered randomized trials for determining more definitively the effects of a given intervention (e.g., Nosek et al., 2018).

A recent preregistered and well-powered randomized controlled trial ($n = 2239$) found that four brief meditation practices (i.e., body scan, mindful breathing, mindful walking, and loving-kindness) all decreased self-reported state anxiety (a construct theoretically related to the negative affect component of subjective wellbeing) compared with a control condition (Sparacio et al., 2024). While this demonstrates that brief meditation practices can be effective, any differences in effectiveness between practices or across individuals are likely to be subtle, making them difficult to detect without sufficiently large samples and appropriate analytical approaches.

By using sufficiently large and demographically heterogeneous samples, it becomes feasible to detect between-practice differences and to support inferences about which individuals may benefit most from a given practice, even if the overall differences are small. For instance, in a preregistered but not well-powered, randomized controlled trial comparing different meditation practices in the context of exposure to a stressor (cold pressor task), a brief gratitude practice improved positive affect compared to a brief mindful breathing practice (Hirshberg et al., 2018). These preliminary findings suggest that certain effects may vary across meditation practices, but it is unclear for whom this is the case. It may be that individual predictors provide relatively little information to inform this matching, as suggested by a recent individual patient data meta-analysis that failed to find evidence that demographic factors, psychological distress, or trait mindfulness moderated the effects of meditation on psychological distress (Galante et al., 2023). However, when multiple predictors are examined collectively using methods such as machine learning, they may provide signals that individual-predictor analyses miss.

Machine learning refers to a wide range of statistical approaches designed to optimize prediction in new individuals by learning patterns from data, often using procedures such as cross-validation to reduce overfitting and evaluate generalizability. In contrast to earlier treatment-matching studies, which typically relied on hypothesis-driven regression models testing one or a small number of baseline patient characteristic × treatment interactions, machine learning approaches can simultaneously incorporate many baseline variables, accommodate multicollinearity, and, in some approaches, capture nonlinear and higher-order interactions (e.g., in decision-tree based approaches; Chekroud et al., 2021). This makes machine learning particularly well suited for estimating individualized treatment effects when differential response is likely to reflect multivariable patterns rather than a single moderator (Webb et al., 2021, 2022a; 2022b, 2026).

Recent studies have utilized machine learning approaches to successfully predict treatment outcomes of psychotherapy, pharmacotherapy, and other mental health interventions (Chekroud et al., 2021; Deisenhofer et al., 2024; Schwartz et al., 2025; Webb et al., 2026). While there have been studies using machine learning approaches to predict which individuals are most likely to benefit from a digital meditation-based wellbeing training versus a control condition (Webb, Hirshberg, et al., 2022, Webb, Hirshberg, et al., 2022; see also Webb, 2021), no study to date has used machine learning approaches to predict which individuals are most likely to benefit from various types of meditation practice. Such a study could potentially optimize the effectiveness of meditation interventions by tailoring meditation practice recommendations to individual profiles.

Using nationally representative samples of the US and UK adult populations with regards to age, sex, and ethnicity, we conducted a randomized controlled trial to investigate the effects of three brief meditation practices that are core components of widely used meditation apps (i.e., mindfulness, self-compassion, and gratitude) relative to a control condition in which participants were asked to use the internet in ways they typically do – an ecologically valid comparison for internet-delivered meditation practices. Our primary hypotheses related to differences by group were that the meditation conditions, compared to the control condition, would show (1a) higher positive affect post-intervention and (1b) lower negative affect post-intervention. While prior studies have shown that brief practices can produce relevant state-level changes (e.g., Palmer et al., 2023), these studies have generally been small and methodologically varied. Given that each meditation practice in this study has a clear theoretical target, our secondary hypotheses were correspondingly that (2a) the mindfulness condition, compared to the other meditation conditions, would show higher state mindfulness post-intervention; (2b) the self-compassion condition, compared to the other meditation conditions, would show higher state self-compassion post-intervention; (2c) the gratitude condition, compared to the other meditation conditions, would show higher state gratitude post-intervention. In exploratory analyses, we examined the effect of intervention dose (five-versus 10-min practices), the effect of speaker gender and ethnicity, and the effect of matching between speaker gender and ethnicity with participants' gender and ethnicity. Our primary hypotheses related to prediction of treatment response were that (3a) pre-intervention characteristics would predict response to the interventions and (3b) participants receiving their algorithm-indicated intervention would show better treatment response than participants treated with a non-indicated intervention.

2. Methods

2.1. Participants

Data were collected between November and December 2023. The study was conducted with US and UK participants (≥18 years old) who were recruited through Prolific Academic (https://app.prolific.co). Beyond an age cut off, no other inclusion or exclusion criteria were applied. Using the representativeness function on the platform (which were only available for US and UK samples), the samples were stratified across three demographics (i.e., age, sex, and ethnicity). Participants were paid 3.2 USD (2.5 GBP) for completing the study, which took approximately 25 min to complete. Informed consent was obtained digitally from participants. Our target sample size was at least 4000 (i.e., ≥1000 per intervention type) participants, which would provide 80% power to detect between-group effects of Cohen's $d \geq 0.10$ for hypotheses 1a and 1b. This sample size also aligns with recommendations for machine learning predictions in digital mental health interventions and exceeds the minimum sample size requirements for treatment selection modeling (Luedtke et al., 2019; Zantvoort et al., 2024). Study procedures were determined to be exempt from review by the Institutional

Review Board at the University of Wisconsin-Madison given researchers had no access to identifying information from participants. The pre-registrations for this study are available on ClinicalTrials.gov (NCT06165497) and the Open Science Framework (https://osf.io/94hks and https://osf.io/fhc6s). Data and analysis code are available at this link: https://osf.io/ap4cj/files/osfstorage.

2.2. Procedure

In the first part of the study (i.e., pre-intervention stage), participants completed a set of baseline items assessing sociodemographic characteristics (e.g., age, gender), questionnaires related to meditation (e.g., perceived barriers to meditation practice; Hunt et al., 2020), personality (Gosling et al., 2003), and baseline psychosocial characteristics (e.g., depressive symptoms, loneliness; Brose et al., 2020; Cohen et al., 1983; Cyranowski et al., 2013; Dweck, 2006; Pilkonis et al., 2011), as well as state affect (Thompson, 2007), trait mindfulness (Baer et al., 2008), trait self-compassion (Raes et al., 2011), and trait gratitude (McCullough et al., 2002). Participants were also presented with an attention check ("I have been randomly selecting responses on this survey"). To allow efficient evaluation of both primary and exploratory research questions, we used a 4 x 2 x 2 factorial randomized trial design (Collins et al., 2014) that crossed intervention type by intervention length by speaker voice. As the control condition did not include a speaker voice type, there were ultimately 14 unique conditions. To retain balance necessary for optimal efficiency (Collins et al., 2014), the sample sizes in the control condition (i.e., using the internet in ways the participants typically do) were twice as large as in the meditation conditions. After completing the initial survey items, participants were randomized using the randomization feature in Qualtrics into one of 14 conditions using a 4 (intervention type [mindfulness, self-compassion, gratitude, control]) x 2 (intervention length [5 min, 10 min]) x 2 (speaker voice [male, female]; as noted, this factor did not apply to the control condition) factorial design: 1) mindfulness, 5 min, male voice; 2) mindfulness, 10 min, male voice; 3) mindfulness, 5 min, female voice; 4) mindfulness, 10 min, female voice; 5) self-compassion, 5 min, male voice; 6) self-compassion, 10 min, male voice; 7) self-compassion, 5 min, female voice; 8) self-compassion, 10 min, female voice; 9) gratitude, 5 min, male voice; 10) gratitude, 10 min, male voice; 11) gratitude, 5 min, female voice; 12) gratitude, 10 min, female voice; 13) control condition, 5 min; 14) control condition, 10 min. As is standard for analysis of factorial experiments (Collins et al., 2014), we collapsed across shared features when examining effects of intervention type, intervention length, and speaker voice.

In the second part of the study (i.e., post-intervention stage), participants completed a manipulation check to ensure that they followed the instructions in the condition to which they were assigned. Specifically, participants were asked to indicate what they were instructed to do (i.e., "pay attention to what was happening in the present moment"; "generate feelings of kindness"; "generate feelings of gratitude"; "use the internet in ways you typically do"). After the manipulation check, participants were assessed on the primary (i.e., state affect; Thompson, 2007) and secondary (i.e., state mindfulness, state self-compassion, and state gratitude; McCullough et al., 2002; Neff et al., 2021; Tanay & Bernstein, 2013) outcomes of the study (see Table S1 for reliability estimates of these outcomes; Table S2 for single survey items; and Table S3 for all scales and subscales), with each scored by averaging item responses. All pre-intervention assessments, the intervention, and post-intervention assessments were completed in a single session. Although participants were also invited to complete a one-week follow-up assessment, these data are reserved for a separate manuscript focused on the durability of effects, given the distinct scope of that question.

Positive affect and negative affect (i.e., the affective components of subjective wellbeing) were the primary outcomes. While the secondary outcomes (i.e., state mindfulness, state self-compassion, and state gratitude) are not facets of subjective wellbeing, they were included to assess practice-specific state changes evoked by each meditation condition. All primary and secondary outcome measures have prior evidence of construct validity in adult samples (McCullough et al., 2002; Neff et al., 2021; Tanay & Bernstein, 2013; Thompson, 2007). Other variables that are theoretically related to negative affect and yet are not synonymous with subjective wellbeing (e.g., depressive symptoms, perceived stress, loneliness) were included as predictors in the machine learning analyses.

3. Materials

3.1. Meditation conditions

Participants randomized to any of the meditation conditions were asked to listen to an audio recording for five or 10 min. The meditation interventions were recordings taken from the Healthy Minds Program app (https://hminnovations.org/meditation-app). Prior research on the Healthy Minds Program has demonstrated it produces benefits on measures of psychological distress and subjective wellbeing (Goldberg et al., 2020; Hirshberg et al., 2022). In the mindfulness condition, participants were instructed to pay attention to what was happening in the present moment with an attitude of non-judgment; in the self-compassion condition, participants were instructed to generate feelings of kindness and friendliness toward themselves; in the gratitude condition, participants were instructed to generate feelings of gratitude and appreciation. Participants pressed a button to start the meditation, which started a timer visible to the participants. Once the designated duration of each meditation condition (i.e., five or 10 min) had elapsed, the button to proceed to the next part of the study was activated, which ensured that participants were required to complete the full meditation session before advancing to the post-intervention survey items.

3.2. Control condition

Participants randomized to the control condition were asked to spend five or 10 min using the internet in ways that they typically do. Mirroring the meditation conditions, participants pressed a button to start a timer and once the designated duration of the control condition (i.e., five or 10 min) had elapsed, the button to proceed to the next part of the study was activated.

4. Analysis plan

4.1. Differences by group assignment

To test our primary hypotheses, we used linear regression models to assess whether there were: (1a) differences in positive affect post-intervention for the meditation conditions versus the control condition, with pre-intervention positive affect included as a covariate; and (1b) differences in negative affect post-intervention for the meditation conditions versus the control condition, with pre-intervention negative affect included as a covariate. All statistical tests were two-tailed ($\alpha = .05$), consistent with the preregistered power analyses. We applied a Benjamini-Hochberg false discovery rate (FDR; Benjamini & Hochberg, 1995) correction for the two primary outcomes. We conducted three preregistered sensitivity analyses for the two primary outcomes. These included running our analyses: (i) with those who failed the manipulation check excluded; (ii) with outliers excluded (i.e., scores three or more SDs from the mean, Cook's distance > 1; Stevens, 1984); and (iii) with missing post-intervention data imputed using multiple imputation.

To test our secondary hypotheses, we used linear regression models to assess whether there were: (2a) differences in state mindfulness scores post-intervention between mindfulness intervention versus other meditation interventions; (2b) differences in state self-compassion scores post-intervention between self-compassion intervention versus other meditation interventions; and (2c) differences in state gratitude scores post-intervention between gratitude intervention versus other meditation interventions. We applied the same correction for multiple comparison (FDR) and the same sensitivity analyses for the three secondary outcomes.

As specified in our preregistration, in exploratory analyses we assessed the impact of intervention length (i.e., 5 min, 10 min), impact of speaker voice (i.e., male, female; non-Latinx White speaker, Latinx speaker), impact of speaker voice gender matching, and impact of speaker voice ethnicity matching. To examine the impact of intervention length and speaker voice, we used the linear regression models specified for the primary outcomes analysis. For the impact of speaker voice gender matching and speaker voice ethnicity matching, we added an interaction between group (e.g., male voice) and demographics (e.g., female gender) to the linear regression models specified for the primary outcomes analysis. While these exploratory analyses were listed in the preregistration, the specific analytic methods used were not preregistered and were chosen during analysis to match the structure of the primary analyses.

4.2. Machine learning prediction of intervention response

In addition to the above preregistered analyses, we created a second registration with two primary machine learning analyses. This was registered after initial analysis of the intervention effects, but prior to any analyses predicting response to interventions. The first prediction analysis (3a) was designed to investigate whether participant pre-intervention characteristics could predict benefit from these interventions. We examined a variety of pre-treatment characteristics (see Tables S2 and S3). To achieve this, we first split the data into a training set (70%) and a holdout test set (30%). The dataset was divided into training and holdout test sets using a temporal split, such that the first 70% of participants enrolled in the study were used to develop the models and the final 30% were reserved for evaluation. Separating model development and evaluation helps reduce the risk of overfitting—that is, when a model captures patterns that are specific to the development dataset rather than generalizable relationships, leading to overly optimistic estimates of predictive performance when the model is evaluated on the same data used for estimation. In a temporal split, model performance is evaluated in participants enrolled later in time, approximating prospective application of the algorithm to new individuals (Steyerberg & Harrell Jr, 2015; Webb et al., 2021, 2022a; 2022b, 2026). We had planned to impute missing data among predictor variables (separately for the training and holdout sample) but given that there was no missingness among predictor variables (as completing these measures was required for randomization), imputation was not needed. Within the training sample, and for each of the four intervention types (i.e., mindfulness, self-compassion, gratitude, and control) separately, we compared the predictive performance of two different modeling approaches. In the preregistration, we stated that we would compare several different machine learning approaches (e.g., random forest, XGBoost, support vector regression, elastic net, ridge regression, and lasso regression) and more conventional approaches (e.g., ordinary least squares linear regression) via nested cross-validation in the training sample. However, in the manuscript, these individual algorithms were implemented within a Super Learner ensemble framework rather than compared separately, as the ensemble approach is designed to optimize predictive performance by combining algorithms via cross-validated weighting. We additionally included a Causal Forest approach (which we were not aware of at the time of preregistration), given its specific design for estimating heterogeneous treatment effects and its ability to flexibly capture nonlinear relationships and interactions (Webb et al., 2026). First, for the Super Learner approach, we included in the ensemble the following machine learning algorithms: (a) five penalized regressions: elastic net with mixing parameter alpha (α) set to either .25, .50 or .75, as well as ridge and lasso (least absolute shrinkage and selection operator) regression (glmnet package; Friedman et al., 2010), (b) two decision-tree based algorithms: RF (randomForest package; Breiman, 2001) and XGBoost (xgboost package; Chen & Guestrin, 2016), (c) support vector regression (with a linear kernel parameter; e1071 package; Steinwart & Christmann, 2008) and (d) two spline regressions: adaptive splines (Friedman, 1991) and adaptive polynomial splines (earth and polspline packages; Stone, Hansen, Kooperberg, & Truong, 1997). We also applied conventional ordinary least square linear regression. The final Super Learner model (one for each of the four groups) represented a linear weighted combination of the above 11 algorithms with weights selected by minimizing cross-validated mean squared error for the predictions (nestcv.SuperLearner function from the R package nestedcv; Lewis et al., 2023).

The Causal Forest approach is an extension of random forests tailored to causal inference with specific focus on estimating heterogeneous treatment effects (i.e., treatment effects as a function of baseline profiles; Wager & Athey, 2018). It modifies the splitting rule of standard regression trees such that samples belonging to the same leaf node can be viewed as generated from a randomized trial and unbiased estimates of the treatment effects can be obtained through sample averages within leaf nodes (Athey & Imbens, 2016). We applied the multi-arm version of Causal Forest where more than two treatments are considered at the same time (Nie & Wager, 2021). The analysis was conducted using multi_arm_causal_forest grf from the grf package (Athey et al., 2019). We used 5000 causal trees and Super Learner for estimating the expected responses marginalized over treatment. All other parameters were kept at their default settings. For stability, the Causal Forest model was repeated 100 times and predictions averaged across iterations. We ultimately conducted machine learning analyses only on the residualized change score itself (i.e., from linear models with post-intervention scores controlling for pre-intervention scores) and did not conduct models predicting post-intervention scores that included pre-intervention scores as covariates. This approach was adopted to simplify the prediction framework while retaining statistical control for baseline symptom levels. Residualized change scores are mathematically equivalent to regression models that predict post-intervention outcomes while controlling for baseline values and are commonly used to adjust for baseline differences when modeling treatment response. Thus, this approach preserved the substantive intent of the preregistered analysis while providing a single outcome variable reflecting baseline-adjusted change. Continuous predictor variables were z-scored (mean = 1 and SD = 1) and binary predictor variables were coded as −0.5 and 0.5. Categorical variables with >2 levels (marital status) were one-hot encoded.

The Super Learner and Causal Forest models were selected to predict outcomes for the participants in the holdout sample (i.e., those participants that the predictive model had not yet "seen"). As a result of the above steps, every participant in the holdout sample had a predicted outcome (residualized change in positive affect/negative affect) score for each of the four intervention type conditions (i.e., mindfulness, self-compassion, gratitude, and control). For each individual, we subtracted their predicted outcome score for the intervention for which they were expected to have the best outcome (i.e., their "indicated" intervention) minus the predicted outcome scores for the other three conditions, and computed the mean of these difference scores. Thus, this Predicted Intervention Benefit (PIB) score reflected the expected magnitude of the benefit an individual was predicted to derive from their indicated intervention relative to the other interventions, on average. Thus, a large positive PIB score indicates that a given individual is expected to derive a relatively large benefit from one specific intervention.

Using ordinary least squares regression models, the second analysis (3b) was designed to test whether participants in the holdout sample who were randomly assigned to their indicated intervention had significantly better outcomes relative to those randomly assigned to their non-indicated treatment and whether PIB scores moderated group differences in outcome via a PIB score x condition received (indicated versus non-indicated) interaction. Observed outcome (i.e., residualized change scores) served as the dependent variable. We screened for univariate (3 SD or greater) and multivariate outliers (Cook's distance > 1) and ran the models with and without outliers.

5. Results

5.1. Descriptive statistics

5260 participants started the study and completed the attention check, of which 155 were excluded from further participation due to attention check failure. 5049 participants completed all survey items up until randomization and were randomized (mindfulness condition = 1262; self-compassion condition = 1268; gratitude condition = 1260; control condition = 1259. 4908 participants completed any post-intervention items (Fig. 1.). The mean age of the participants who started the study was 45 and the majority of them reported being female (50.9%), heterosexual (84.4%), White (80.2%), employed (70.7%), and not at all religious (51.5%; see Tables S2–S4 for information about survey items, questionnaires, and sample characteristics). The meditation interventions did not differ from the control condition on positive or negative affect at pre-intervention ($ps \geq .334$). Similarly, the meditation interventions did not differ from one another on positive or negative affect at pre-intervention ($ps \geq .127$). The majority of participants (78.0%) passed the manipulation check. Participants in the self-compassion and gratitude condition were less likely to pass the manipulation check (68.2% and 53.2%, respectively) than those in the mindfulness (92.5%) and control (98.3%) conditions. This may have been due to participants in the self-compassion and gratitude conditions having difficulty distinguishing between generating feelings of kindness and generating feelings of gratitude.

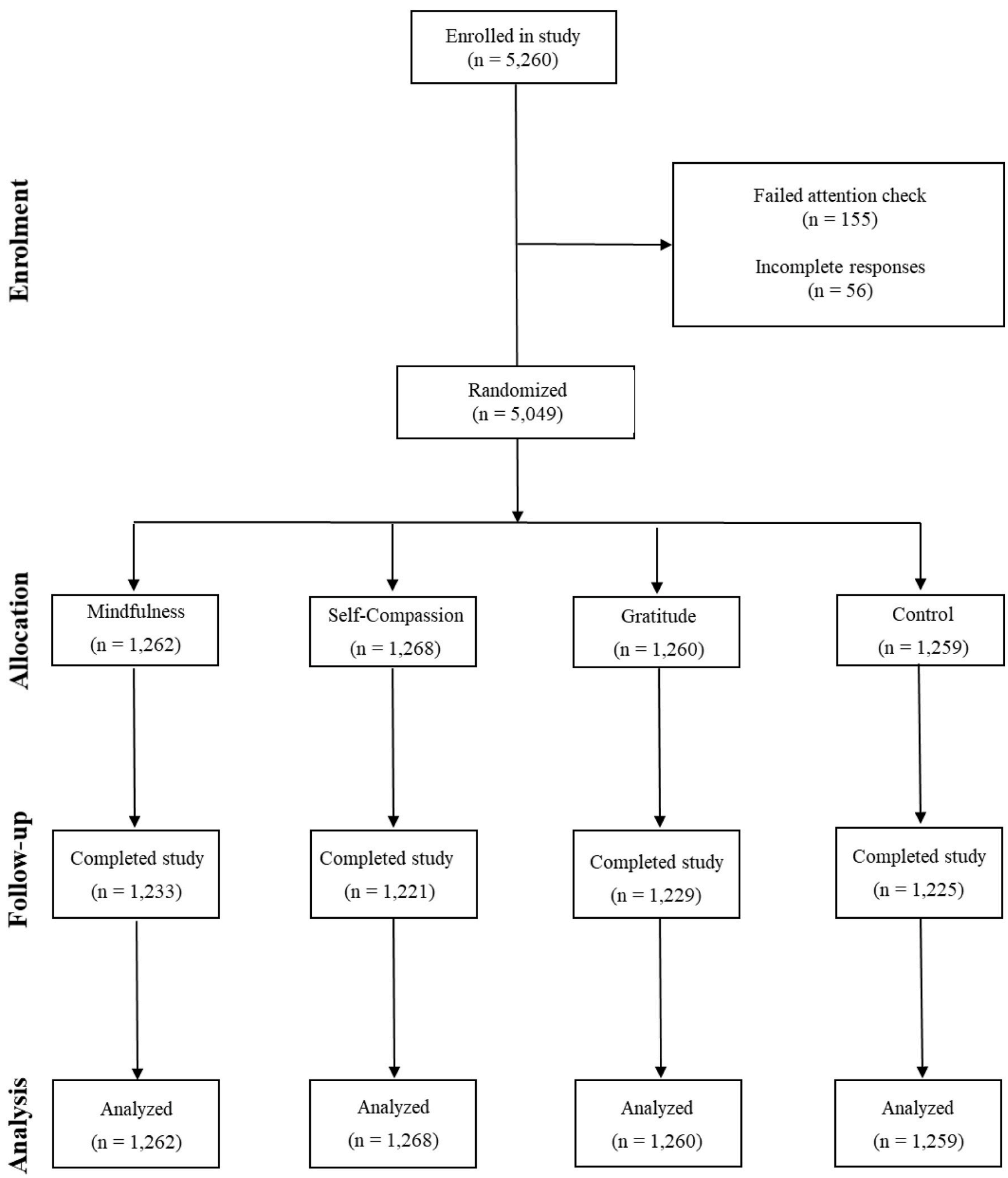

Fig. 1. CONSORT flow diagram.

5.2. Inferential statistics

Primary Outcomes. The sample overall increased in positive affect and decreased in negative affect from pre-to post-intervention (Cohen's $ds = 0.08$ and -0.27, $t(4907) = 9.73$ and -28.69, $ps < .001$, respectively; Table 1). Increases in positive affect did not differ between the meditation conditions and the control condition (within-group $ds = 0.09$ versus 0.06, respectively, β (standardized coefficient) $= 0.011$, $t(4905) = 1.39$, $p_{FDR} = 0.165$; see Table 1 and Table 2). Decreases in negative affect were of larger magnitude for the meditation conditions than for the control condition ($ds = -0.30$ versus -0.20, respectively, $\beta = -0.054$, $t(4905) = -5.72$, $p_{FDR} < .001$; see Table 1 and Table 2).

Secondary Outcomes. Hypothesis 2a was not supported. Unexpectedly, state mindfulness of mental events was lower post-intervention in the mindfulness condition than in the other meditation conditions ($\beta = -0.077$, $t(3681) = -4.66$, $p_{FDR} < .001$; Table 2) – a significant difference in the direction opposite to our prediction. The mindfulness condition did not differ from the other meditation conditions on state mindfulness of physical sensations ($\beta = -0.005$, $t(3681) = -0.29$, $p_{FDR} = 0.773$; Table 2). Consistent with our hypothesis, the self-compassion condition was associated with higher state self-compassion post-intervention relative to the other meditation conditions ($\beta = 0.047$, $t(3681) = 2.84$, $p_{FDR} = 0.009$; Table 2). Similarly, the gratitude condition was associated with higher state gratitude post-intervention relative to the other meditation conditions ($\beta = 0.045$, $t(3681) = 2.71$, $p_{FDR} = 0.009$; Table 2; see Table S5 for descriptive statistics on secondary outcomes).

Sensitivity Analyses. Significance tests were unchanged and effect sizes were very similar across sensitivity analyses that restricted the sample to those who passed the manipulation check, with outliers removed, and with missing values replaced using multiple imputation (see Table S6 for sensitivity analyses).

Exploratory Analyses. ***Dose Effects.*** Collapsing across the meditation intervention types, 5-min interventions were associated with larger increases in positive affect than 10-min interventions ($ds = 0.11$ versus 0.06, $\beta = 0.026$, $t(3680) = 2.81$, $p_{FDR} = 0.010$; see Tables S7–S10). There were no effects of dose on negative affect ($ds = -0.28$ and -0.31, for 5-min and 10-min interventions, respectively, $\beta = -0.003$, $t(3680) = -0.29$, $p_{FDR} = 0.771$; see Tables S7–S10).

Speaker Gender and Ethnicity Effects. Collapsing across the meditation interventions, there was no effect of speaker gender and ethnicity on positive or negative affect ($\beta s = -0.004$ and -0.015, $t(3680) = -0.39$ and -1.29, $p_{FDR}s = 0.693$ and .391, respectively, with Latinx female as the reference group; see Tables S7–S10).

Speaker Gender and Ethnicity Matching. Collapsing across the meditation interventions, there was no interaction between participant gender and speaker gender on positive or negative affect (interaction $bs = -0.003$ and 0.015, $t(3678) = -0.09$ and 0.62, $p_{FDR}s = 0.927$, respectively; see Tables S7–S10). Similarly, there was no interaction between ethnic minority participant identity and speaker ethnic identity on positive or negative affect (interaction $bs = -0.027$ and 0.005, $t(3678) = -0.57$ and 0.15, $p_{FDR}s = 0.880$, respectively; see Tables S7–S10).

5.3. Prediction models

Change in Positive Affect. ***Predicted Outcomes.*** The Super Learner model predicted that 29.6% of participants in the holdout sample would experience the greatest increases in positive affect receiving self-compassion, compared to 26.7% for control condition, 22.5% for mindfulness, and 21.2% for gratitude. The Causal Forest model predicted that 44.5% of participants would benefit most from self-compassion, 29.6% from gratitude, 14.3% from control condition, and 11.6% from mindfulness. The top 10 most important baseline features (i.e., variables that most strongly moderated group differences in outcome) from the Causal Forest model (averaged across the 100 iterations) are presented in Fig. S1.

Evaluating Model Recommendations. Individuals in the holdout sample who were randomly assigned to their Super Learner-indicated intervention had significantly better outcomes (i.e., greater increases in positive affect) than those who received a non-indicated intervention ($b = 0.074$, $t(1372) = 2.05$, $p = .040$, $d = 0.13$) (Fig. 2, left panel). This effect was no longer significant when univariate outliers were removed ($b = 0.057$, $t(1353) = 1.71$, $p = .087$). No multivariate outliers were detected. For the Causal Forest algorithm, those receiving the Causal Forest-indicated intervention did not have significantly better outcomes relative to those who received a non-indicated intervention ($b = 0.053$, $t(1372) = 1.48$, $p = .140$, $d = 0.09$) (Fig. 3, left panel). Results were similar with univariate outliers removed ($b = 0.049$, $t(1353) = 1.51$, $p = .131$). No multivariate outliers were detected.

For the Super Learner-derived intervention recommendations, there was no PIB x Got Indicated (i.e., did or did not receive their Super Learner-indicated intervention) interaction ($b = -0.19$, $t(1370) = -0.36$, $p = .718$; Fig. 2, right panel). Results were similar with univariate outliers removed ($b = -0.15$, $t(1339) = -0.29$, $p = .770$). No multivariate outliers were detected. For the Causal Forest model, there was also no PIB x Got Indicated interaction ($b = 1.17$, $t(1370) = 0.34$, $p = .736$; Fig. 3, right panel). Results were similar with univariate outliers removed ($b = 0.73$, $t(1351) = 0.23$, $p = .817$). No multivariate outliers were detected. Given the PIB x Got Indicated interaction was not significant for either the Super Learner or Causal Forest models, we did not probe this interaction further.

Secondary Analyses. One question is to what extent these algorithms can identify subgroups of individuals best suited to each of the three

Table 1
Pre- and post-intervention descriptive statistics.

Sample	Outcome	Pre-intervention			Post-intervention			Change			Cohen's d	d_{LB}	d_{UB}
		n	Mean	SD	n	Mean	SD	n	Mean	SD			
Full sample	Positive affect	5049	2.95	0.95	4908	3.03	0.99	4908	0.08	0.56	0.08	0.06	0.10
Full sample	Negative affect	5049	1.45	0.70	4908	1.26	0.56	4908	−0.19	0.47	−0.27	−0.28	−0.26
Meditation groups	Positive affect	3790	2.94	0.95	3683	3.03	0.99	3683	0.08	0.58	0.09	0.07	0.11
Meditation groups	Negative affect	3790	1.45	0.70	3683	1.24	0.53	3683	−0.21	0.48	−0.30	−0.32	−0.28
Control	Positive affect	1259	2.97	0.97	1225	3.02	0.99	1225	0.06	0.50	0.06	0.03	0.09
Control	Negative affect	1259	1.45	0.70	1225	1.31	0.61	1225	−0.14	0.43	−0.20	−0.22	−0.18
Mindfulness	Positive affect	1262	2.94	0.94	1233	3.01	0.99	1233	0.07	0.6	0.08	0.05	0.11
Mindfulness	Negative affect	1262	1.46	0.69	1233	1.23	0.51	1233	−0.23	0.49	−0.33	−0.36	−0.30
Self-compassion	Positive affect	1268	2.96	0.95	1221	3.04	0.97	1221	0.08	0.57	0.08	0.05	0.11
Self-compassion	Negative affect	1268	1.43	0.69	1221	1.23	0.50	1221	−0.19	0.45	−0.27	−0.30	−0.24
Gratitude	Positive affect	1260	2.93	0.95	1229	3.03	0.99	1229	0.10	0.57	0.11	0.08	0.14
Gratitude	Negative affect	1260	1.47	0.73	1229	1.27	0.59	1229	−0.2	0.49	−0.29	−0.32	−0.26

Note: Positive and negative affect = PANAS; SD = standard deviation. Cohen's *d* calculated based on the pre-post change divided by the pre-intervention standard deviation (Morris & DeShon, 2002); d = Cohen's *d*; 95% confidence interval LB = lower bound; 95% confidence interval UB = upper bound.

Table 2
Between-group differences on post-intervention primary and secondary outcomes.

Model	Outcome	β	β_{LB}	β_{UB}	*t*	*df*	*p*	p_{FDR}
Meditation versus control	Positive affect	0.011	−0.005	0.026	1.39	4905	.165	.165
Meditation versus control	Negative affect	−0.054	−0.073	−0.036	−5.72	4905	<.001	<.001
Mindfulness versus other meditations	Mindfulness - mental	−0.077	−0.109	−0.044	−4.66	3681	<.001	<.001
Mindfulness versus other meditations	Mindfulness - physical	−0.005	−0.037	0.028	−0.29	3681	.773	.773
Self-compassion versus other meditations	Self-compassion	0.047	0.014	0.079	2.84	3681	.005	.009
Gratitude versus other meditations	Gratitude	0.045	0.012	0.077	2.71	3681	.007	.009

Note: Positive and negative affect = PANAS; Mindfulness = State Mindfulness Scale; Self-compassion = State Self-Compassion Scale; Gratitude = Gratitude Adjective Checklist – 3. β = standardized regression coefficient (scaled predictor and scaled outcome); 95% confidence interval LB = lower bound; 95% confidence interval UB = upper bound; *t* = *t*-statistic; *df* = degrees of freedom; p_{FDR} = false discovery rate-corrected *p*-value.

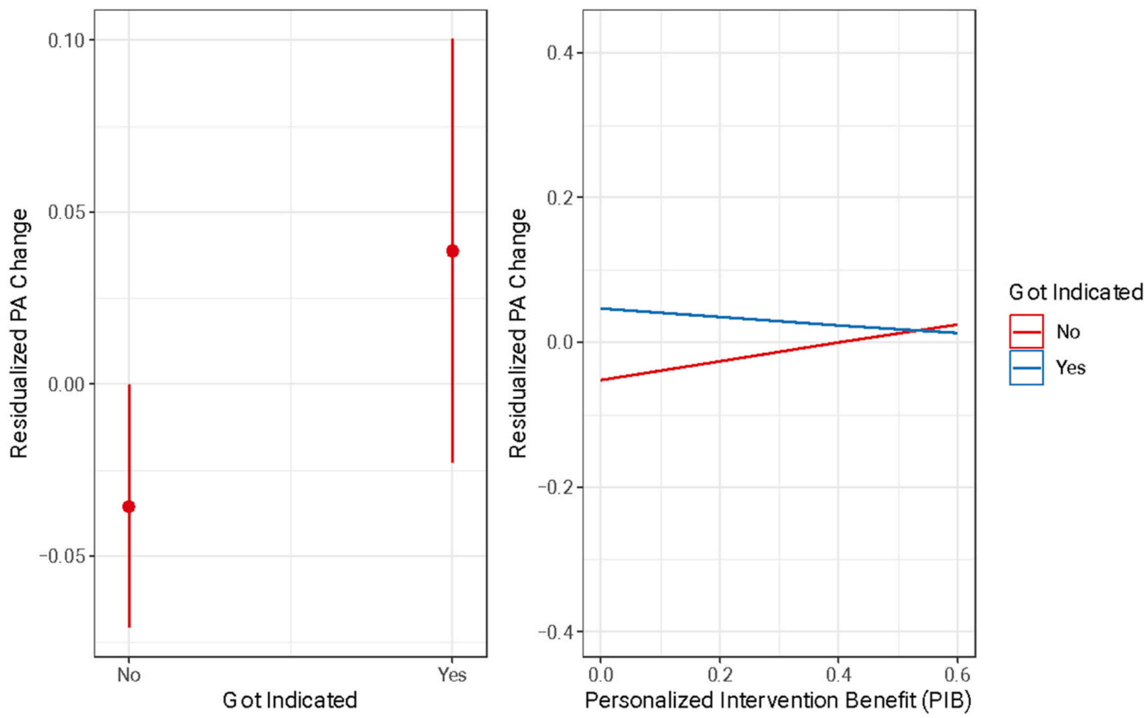

Fig. 2. Super Learner (SL) results: Residualized positive affect (PA) change as a function of (1) whether holdout participants received their SL-indicated versus non-indicated intervention (left panel), or (2) Personalized Intervention Benefit (PIB) continuous score (right panel).

meditation interventions (i.e., mindfulness, self-compassion, and gratitude) versus merely distinguishing between subgroups of individuals who benefit from any of the three meditation conditions versus the control condition. To address this, we re-ran the significant Super Learner analysis removing all participants who received the control condition in the holdout sample. Individuals assigned to their Super Learner-indicated meditation intervention did not differ significantly from those receiving a non-indicated intervention (b = 0.08, t(1016) = 1.77, p = .078, d = 0.13).

Change in Negative Affect. ***Predicted Outcomes.*** The Super Learner model predicted that 39.3% of participants in the holdout sample would experience the greatest reductions in negative affect receiving self-compassion, compared to 34.5% for mindfulness, 21.3% for gratitude, and 4.8% for control condition. The Causal Forest model predicted that 65.7% of participants would benefit most from mindfulness, 20.1% from self-compassion and 14.3% from gratitude. For the Causal Forest model, no participants were predicted to have the best outcome to the control condition. The top 10 most important baseline features (i.e., variables that most strongly moderated group differences in outcome) from the Causal Forest model (averaged across the 100 iterations) are presented in Fig. S2.

Evaluating Model Recommendations. Individuals in the holdout sample who were randomly assigned to their Super Learner-indicated intervention did not have significantly better outcomes (i.e., greater reductions in negative affect) than those who received a non-indicated intervention (b = −0.011, t(1372) = −0.49, p = .627, d = 0.03; Fig. 4, left panel). Results were similar with univariate outliers removed (b = −0.022, t(1341) = −1.23, p = .220). No multivariate outliers were detected. For the Causal Forest algorithm, those receiving the Causal Forest-indicated intervention had significantly better outcomes relative to those who received a non-indicated intervention (b = −0.062, t(1372) = −2.58, p = .010, d = 0.16; Fig. 5, left panel). Results were similar with univariate outliers removed (b = −0.052, t(1341) = −2.94, p = .003). No multivariate outliers were detected.

For the Super Learner-derived intervention recommendations, the PIB x Got Indicated interaction was not significant (b = −0.34, t(1370) = −1.66, p = .098; Fig. 4, right panel). However, the PIB x Got Indicated interaction was significant when univariate outliers were removed (b = −0.56, t(1318) = −3.16, p = .002). No multivariate outliers were detected. For the Causal Forest model, the PIB x Got Indicated interaction was significant (b = −2.19, t(1370) = −4.94, p < .001), which revealed that PIB scores moderated group differences in outcomes between receiving ones indicated versus a non-indicated intervention. Results were similar with univariate outliers removed (b = −1.96, t(1337) = −5.76, p < .001). No multivariate outliers were detected. As shown in Fig. 5 (right panel), larger PIB scores (indicating a

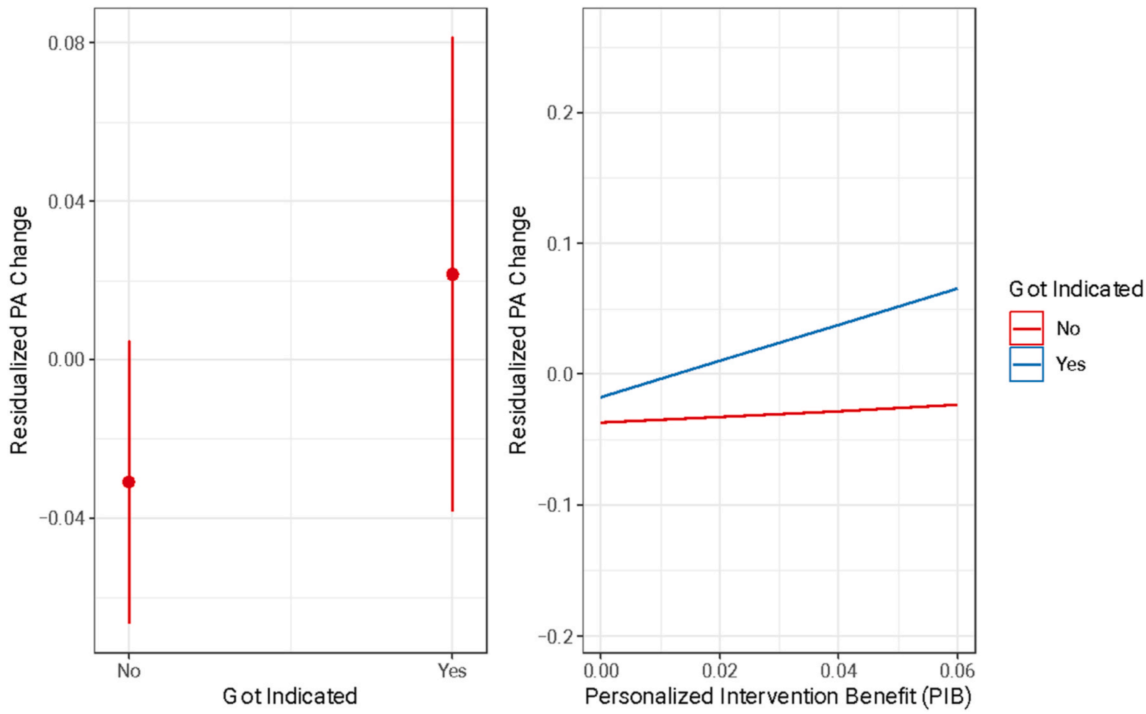

Fig. 3. Causal Forest (CF) results: Residualized positive affect (PA) change as a function of (1) whether holdout participants received their CF-indicated versus non-indicated intervention (left panel), or (2) Personalized Intervention Benefit (PIB) continuous score (right panel).

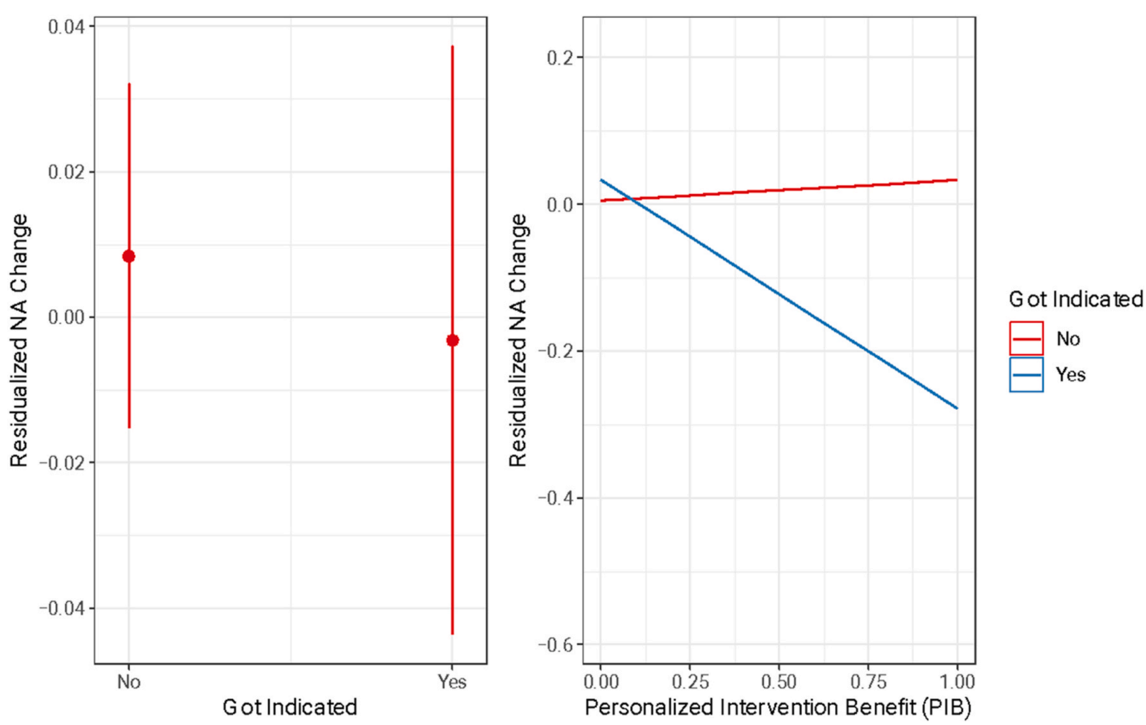

Fig. 4. Super Learner (SL) results: Residualized negative affect (NA) change as a function of (1) whether holdout participants received their SL-indicated versus non-indicated intervention (left panel), or (2) Personalized Intervention Benefit (PIB) continuous score (right panel).

greater predicted benefit from the indicated intervention) were associated with greater actual benefit (i.e., reduction in negative affect) when participants received their Causal Forest-indicated intervention in the holdout sample.

Evaluating Model Recommendations (Larger PIB Scores). Individuals with small PIB scores (near zero) are predicted to have a limited or no advantage from assignment to one intervention. Given this, we compared the effect size for the group difference (received Causal Forest-indicated versus non-indicated intervention) within the holdout subsample among (1) the top 50% and (2) top 25% of PIB scores, and for whom the 100 iterations of the Causal Forest model resulted in 90% or greater consistency in which intervention was predicted to be optimal

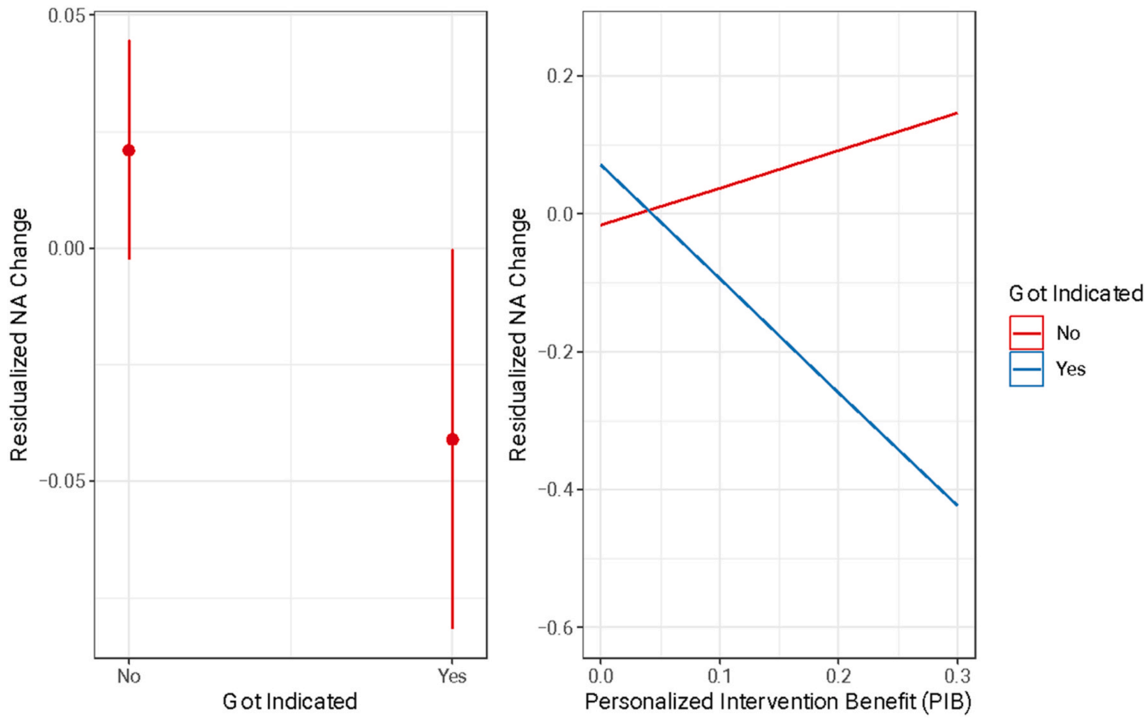

Fig. 5. Causal Forest (CF) results: Residualized negative affect (NA) change as a function of (1) whether holdout participants received their CF-indicated versus non-indicated intervention (left panel), or (2) Personalized Intervention Benefit (PIB) continuous score (right panel).

(n = 708 and n = 355, respectively). As expected, effect sizes for these two groups (d = 0.30 and 0.43, respectively) were larger than in the full sample (d = 0.16; Fig. 6).

Secondary Analyses. Similar to the positive affect models, we re-ran the significant Causal Forest analysis removing all participants who received the control condition in the holdout sample in order to

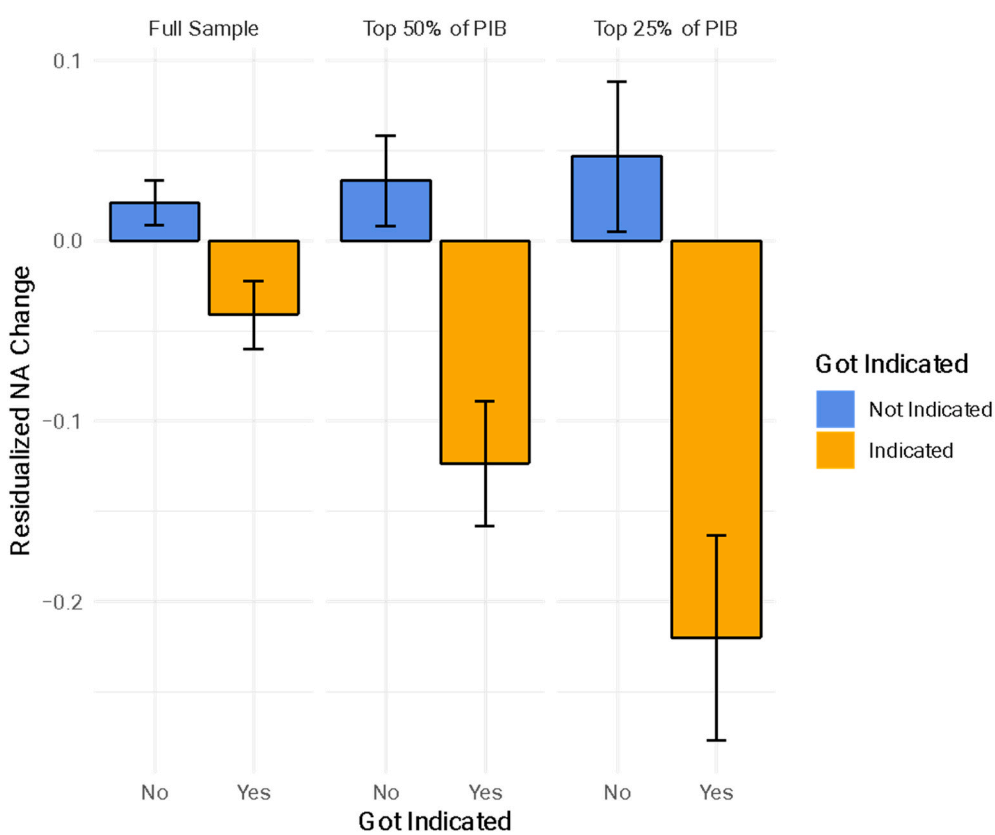

Fig. 6. Mean residualized change in negative affect (NA) for individuals in the holdout sample who did versus did not receive their indicated intervention according to the Causal Forest algorithm at varying subsets of Personalized Intervention Benefit (PIB) scores (error bars indicating ±1 standard error of the mean).

determine to what extent the algorithm could identify subgroups of individuals best suited to each of the three meditation interventions versus merely distinguishing between meditation conditions versus control condition. Results showed that individuals assigned to their Causal Forest-indicated meditation intervention had significantly better outcomes compared to those receiving a non-indicated intervention ($b = -0.06$, $t(1016) = -2.28$, $p = .023$, $d = 0.15$).

6. Discussion

This preregistered and well-powered study investigated the effects of three brief meditation practices on subjective wellbeing, compared with a control condition, as well as whether pre-intervention characteristics predicted response to the different conditions and whether participants who received their algorithm-indicated intervention showed better treatment response than participants treated with a non-indicated intervention. Given these practices form the basis of widely used digital meditation interventions (e.g., meditation apps; Wasil et al., 2020), studying brief practices can help understand the proximal effects of these practices as well as evaluate our ability to predict response to them on immediate outcomes. Results showed no significant difference between the meditation conditions and the control condition on positive affect, but decreases in negative affect were greater for the meditation conditions relative to the control condition. This broadly corresponds with previous research on brief meditation interventions (Schumer et al., 2018). Results also showed preliminary evidence, from one of two analytical approaches, that pre-intervention characteristics can predict preferential response to brief meditation interventions for change in negative affect. This pattern emerged in the Causal Forest analyses but not in the Super Learner analyses, and these findings should therefore be interpreted with caution and replicated before conclusions are drawn.

The primary findings suggest that brief meditation interventions may have potential as accessible, affordable and scalable mental health tools, particularly for reducing negative affect. The control condition provides an ecologically valid comparison – individuals who used five or 10 min listening to a guided meditation showed decreases in negative affect compared to those who used the internet in ways they typically do (which could have, in theory, involved other strategies that could improve subjective wellbeing such as watching humorous videos; Stieger et al., 2023). The findings also underscore the promise of personalized approaches in contemplative science and indicate that mental health outcomes may be enhanced by tailoring interventions based on an individual's pre-intervention characteristics. Even if the effect sizes in this study were small, the interventions being tested are very brief and scalable – implementation and delivery of the study's insights at scale through digital meditation practices could have substantial impact and public health relevance (Funder & Ozer, 2019; Prentice & Miller, 1992), especially considering the relatively low burden and low costs of these types of interventions.

Secondary analyses showed mixed evidence for the unique effects of the three meditation practices. While participants in the self-compassion and gratitude conditions scored higher post-intervention on outcomes related to the respective interventions (i.e., state self-compassion and state gratitude), there was no post-intervention difference across conditions on mindfulness of physical sensations. Notably, scores on state mindfulness of mental events were unexpectedly lower post-intervention in the mindfulness condition – the opposite of our preregistered prediction. One possible explanation for this that will be important to investigate in future studies is that individuals who received the mindfulness intervention were simply more aware of their lack of awareness. This possibility would be consistent with the Dunning-Kruger effect that has documented paradoxical linkages between lower abilities and higher self-assessments for other domains (e.g., humor, grammar, logic; Kruger & Dunning, 1999). This phenomenon has been attributed to a lack of metacognitive abilities and it is possible that the brief mindfulness training delivered here increased metacognitive abilities such that individuals were able to more accurately report on their mental state (i.e., lack of awareness). An alternative explanation is that self-compassion and gratitude practices genuinely increased state mindfulness to a larger extent than explicit mindfulness training, which is certainly possible since these practices also involve regulating attention.

Our factorial design allowed us to conduct exploratory analyses related to intervention length (i.e., dose) and speaker voice type. When collapsing across the meditation intervention types, exploratory analyses showed that the 5-min interventions were associated with larger increases in positive (but not negative) affect relative to 10-min interventions, although neither dose differed significantly from the control condition on positive affect. The impact of dosage in digital meditation training is an active area of debate and it is not currently clear whether higher amounts of practice produce greater benefits (Goldberg et al., 2025). To our knowledge, the current study is the largest experimental test of the impact of two meditation doses (i.e., five- or 10-min practices). In a sample of participants with generally limited prior meditation experience (52.0% of those who completed post-intervention assessments reported $\leq$10 total lifetime hours of meditation practice of the types of meditation assessed; see Table S2 for items), it may very well be that shorter practices are more immediately beneficial than even slightly longer practices. Meditation involves cognitive and affective processes that may initially be perceived as effortful, although these perceptions appear to change toward less effort and more enjoyment with more training (Lumma et al., 2015; Lutz et al., 2015). Large-scale studies manipulating meditation dosage in longer-term training are needed. This is an obvious feature that could easily be personalized through dosage recommendations within digital meditation training.

In contrast to differences across doses, when collapsing across the meditation intervention types, exploratory analyses showed no effect of speaker gender and ethnicity on positive or negative affect, no interaction between participant gender and speaker gender on positive or negative affect, and no interaction between ethnic minority participant identity and speaker ethnic identity on positive or negative affect. These null findings suggest that the gender or ethnicity of meditation audio voices may not have a meaningful impact on subjective wellbeing in brief meditation interventions. Given documented demographic disparities in access to digital meditation interventions (Jiwani et al., 2023), intervention developers may need to move beyond more surface-level adaptations (e.g., speaker demographics) and engage in deeper cultural adaptation such as adapting the intervention content itself (Benish et al., 2011). Although it is possible that we failed to detect the true effect of matching between participant and speaker identities, a power calculation using the 'pwr' package in R (Champely et al., 2020) indicated that we had sufficient power to detect very small effects for this interaction ($f^2 < 0.003$). This was true even in the presence of some degree of imbalance for the ethnicity variable (coded as White and non-White, with 80.2% identifying as White).

While the findings in the current research are promising, it is important to consider certain limitations. First, the outcomes were self-reported, which may have biased the results due to demand characteristics present more in the meditation conditions than the control condition. Second, the sample was stratified to be representative of the respective national populations with regards to age, sex, and ethnicity, but it is possible that the sample was not representative with regards to other sociodemographic characteristics. The findings may therefore not be generalizable to the US or UK adult populations. It is also possible that the findings may not generalize to adult populations in other countries (e.g., those with very different socio-cultural and socio-economic landscapes from the US and UK), which highlights the need for further research in various countries. Notably, research in low- and middle-income countries may be particularly useful, given these contexts may stand to especially benefit from highly scalable and low-cost interventions (Sun et al., 2024). Third, the control condition in this study involved using the internet in ways the participants' typically do, but the participants were not asked how they used the internet during the study nor was this objectively monitored. Although this was chosen as an ecologically valid control for meditation practices also being delivered via the internet, it is likely that the way in which participants used the internet differed in terms of impact on outcomes. It would have been helpful to have monitored how participants used the internet in the control group, to have employed a standardized comparison group with a relaxing activity (e.g., listening to music or an audiobook; Sparacio et al., 2024), or even to have asked participants explicitly to spend time using the internet to relax in any way they wish. Fourth, there was no external validation of the data-driven approach in a separate sample. In addition, the machine learning analyses were only significant for negative, but not for positive, affect. It is also important to acknowledge that there are additional challenges with machine learning approaches in general that further complicates generalizability (e.g., Vela et al., 2022). Fifth, the secondary outcomes were assessed only at post-intervention. This meant that the secondary analyses could not adjust for pre-intervention scores on the corresponding state measures, unlike the primary analyses, which adjusted for pre-intervention positive and negative affect. While randomization makes systematic pre-intervention differences across conditions unlikely, the absence of these baseline covariates precluded the additional statistical precision such adjustment would have provided. Future research on treatment personalization should use more objective measures of subjective well-being (e.g., via behavioral tasks) and more representative samples from various countries in a prospective randomized controlled trial to test the same hypotheses and externally validate the data-driven approach used in this study in a new sample (Lutz et al., 2022).

In conclusion, this study investigated the effects of brief meditation practices on subjective wellbeing and the potential of using pre-intervention characteristics to predict response to brief meditation interventions. Taken together, the findings suggest that brief meditation practices (i.e., mindfulness, self-compassion, and gratitude) modestly reduce negative affect and response to these practices can be predicted using pre-interventions characteristics. Based on the machine learning algorithms from this study (e.g., the significant Causal Forest model predicting change in negative affect), a future study could randomly assign participants to either an algorithm-assigned intervention condition (i.e., allocated to the meditation intervention that predicts optimal response) versus a relevant comparison condition (e.g., randomly allocated to one of the three meditation interventions, or participant choice based on their personal preference; Marcus et al., 2012). Such an implementation study can directly test the benefits of treatment personalization using pre-intervention characteristics.

Funding

OS was supported by Ekhaga Foundation, Olle Engkvist Foundation, and Norrsken Mind. CAW was partially supported by NIMH R01MH116969, NCCIH R01AT011002, the Tommy Fuss Fund and a Young Investigator Grant from the Brain & Behavior Research Foundation. SBG was supported by the National Center for Complementary & Integrative Health award K23AT010879. MJH was partially funded by a National Institute of Mental Health award K01MH130752. This study was funded by grants from the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement 964038, NUS STAR account, and Olle Engkvist Foundation.

CRediT authorship contribution statement

Otto Simonsson: Conceptualization, Funding acquisition, Methodology, Writing – original draft. **Christian A. Webb:** Formal analysis, Funding acquisition, Writing – review & editing. **Mini Ruiz:** Funding acquisition, Writing – review & editing. **Walter Osika:** Writing – review & editing. **Jayanth Narayanan:** Funding acquisition, Writing – review & editing. **Micael Dahlén:** Writing – review & editing. **Wolfgang Lutz:** Formal analysis, Writing – review & editing. **Brian Schwartz:** Formal analysis, Writing – review & editing. **Boyu Ren:** Formal analysis, Writing – review & editing. **Cortland J. Dahl:** Resources, Writing – review & editing. **Richard J. Davidson:** Writing – review & editing. **Matthew J. Hirshberg:** Funding acquisition, Methodology, Writing – review & editing. **Simon B. Goldberg:** Conceptualization, Formal analysis, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interests

OS is a co-founder of embla.ai AB and has once received a small payment from Mindfully Sweden AB for educational content. CAW has received consulting fees from King & Spalding law firm. RJD is the founder, president, and serves on the board of directors for the non-profit organization, Healthy Minds Innovations, Inc.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2026.105141.

Data availability

Data and analysis code are available at this link: https://osf.io/ap4cj/files/osfstorage

References

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences, 113*(27), 7353–7360.
Athey, S., Tibshirani, J., & Wager, S. (2019). *Generalized random forests*.
Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., … Gu, J. (2008). FFMQ-15: 15-item five-facet mindfulness questionnaire. *Assessment, 15*(3), 329–342.
Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology, 81*(4), 722.
Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. *Journal of Counseling Psychology, 58*(3), 279.
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*(1), 289–300.
Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.
Brose, A., Raedt, R. D., & Vanderhasselt, M. A. (2020). Eight items of the ruminative response scale are sufficient to measure weekly within-person variation in rumination. *Current Psychology*, 1–9.
Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., … De Rosario, H. (2020). Pwr: Basic functions for power analysis. *R package version 1.3-0)* [Computer software] https://CRAN.R-project.org/package=pwr.
Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., … Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry, 20*(2), 154–170.
Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).
Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 385–396.
Collins, L. M., Dziak, J. J., Kugler, K. C., & Trail, J. B. (2014). Factorial experiments: Efficient tools for evaluation of intervention components. *American Journal of Preventive Medicine, 47*(4), 498–504.
Cyranowski, J. M., Zill, N., Bode, R., Butt, Z., Kelly, M. A., Pilkonis, P. A., … Cella, D. (2013). Assessing social support, companionship, and distress: National institute of health (NIH) toolbox adult social relationship scales. *Health Psychology, 32*(3), 293.
Deisenhofer, A. K., Barkham, M., Beierl, E. T., Schwartz, B., Aafjes-van Doorn, K., Beevers, C. G., … Cohen, Z. D. (2024). Implementing precision methods in personalizing psychological therapies: Barriers and possible ways forward. *Behaviour Research and Therapy, 172*, Article 104443.
Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542.
Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*(2), 276.
Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random house.
Folk, D., & Dunn, E. (2023). A systematic review of the strength of evidence for the Most commonly recommended happiness strategies in mainstream media. *Nature Human Behaviour, 7*(10), 1697–1707.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Annals of Statistics, 19*(1), 1–67.
Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*, 1–22.
Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in methods and practices in psychological science, 2*(2), 156–168.
Galante, J., Friedrich, C., Dalgleish, T., Jones, P. B., & White, I. R. (2023). Individual participant data systematic review and meta-analysis of randomised controlled trials assessing adult mindfulness-based programmes for mental health promotion in non-clinical settings. *Nature Mental Health, 1*(7), 462–476.
Goldberg, S. B., & Davidson, R. J. (2024). Contemplative science comes of age: Looking backward and forward 20 years after baer (2003). *Clinical Psychology: Science and Practice, 31*(1), 39.
Goldberg, S. B., Imhoff-Smith, T., Bolt, D. M., Wilson-Mendenhall, C. D., Dahl, C. J., Davidson, R. J., & Rosenkranz, M. A. (2020). Testing a multi-component, self-guided, smartphone-based meditation app: Three-armed randomized controlled trial. *JMIR Mental Health, 7*(11), Article e23825.
Goldberg, S. B., Kendall, A. D., Hirshberg, M. J., Dahl, C. J., Nahum-Shani, I., Davidson, R. J., & Bray, B. C. (2025). Is dosage of a meditation app associated with changes in psychological distress? It depends on how you ask. *Clinical Psychological Science, 13*(2), 332–349.
Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*(6), 504–528.
Hirshberg, M. J., Frye, C., Dahl, C. J., Riordan, K. M., Vack, N. J., Sachs, J., … Goldberg, S. B. (2022). 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, 114*(8), 1895.
Hirshberg, M. J., Goldberg, S. B., Schaefer, S. M., Flook, L., Findley, D., & Davidson, R. J. (2018). Divergent effects of brief contemplative practices in response to an acute stressor: A randomized controlled trial of brief breath awareness, loving-kindness, gratitude or an attention control practice. *PLoS One, 13*(12), Article e0207765.
Hunt, C. A., Hoffman, M. A., Mohr, J. J., & Williams, A. L. (2020). Assessing perceived barriers to meditation: The determinants of meditation practice inventory-revised (DMPI-R). *Mindfulness, 11*, 1139–1149.
Ioannidis, J. P. (2005). Why Most published research findings are false. *PLoS Medicine, 2*(8), Article e124.
Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., Gortner, E., & Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64*(2), 295–304.
Jiwani, Z., Tatar, R., Dahl, C. J., Wilson-Mendenhall, C. D., Hirshberg, M. J., Davidson, R. J., & Goldberg, S. B. (2023). Examining equity in access and utilization of a freely available meditation app. *Npj mental health research, 2*(1), 5.
Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121.
Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour, 4*(4), 423–434.
Lewis, M. J., Spiliopoulou, A., Goldmann, K., Pitzalis, C., McKeigue, P., & Barnes, M. R. (2023). Nestedcv: An R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data. *Bioinformatics Advances, 3*(1). vbad048.
Lindsay, E. K., Chin, B., Greco, C. M., Young, S., Brown, K. W., Wright, A. G., … Creswell, J. D. (2018). How mindfulness training promotes positive emotions: Dismantling acceptance skills training in two randomized controlled trials. *Journal of Personality and Social Psychology, 115*(6), 944.
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.
Lumma, A. L., Kok, B. E., & Singer, T. (2015). Is meditation always relaxing? Investigating heart rate, heart rate variability, experienced effort and likeability during training of three types of meditation. *International Journal of Psychophysiology, 97*(1), 38–45.
Lutz, A., Jha, A. P., Dunne, J. D., & Saron, C. D. (2015). Investigating the phenomenological matrix of mindfulness-related practices from a neurocognitive perspective. *American Psychologist, 70*(7), 632.
Lutz, W., Schwartz, B., & Delgadillo, J. (2022). Measurement-based and data-informed psychological therapy. *Annual Review of Clinical Psychology, 18*(1), 71–98.
Marcus, S. M., Stuart, E. A., Wang, P., Shadish, W. R., & Steiner, P. M. (2012). Estimating the causal effect of randomization versus treatment preference in a doubly randomized preference trial. *Psychological Methods, 17*(2), 244.
Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods, 9*(2), 147.
McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology, 82*(1), 112.
Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105.
Neff, K. D., Tóth-Király, I., Knox, M. C., Kuchar, A., & Davidson, O. (2021). The development and validation of the state self-compassion scale (long-and short form). *Mindfulness, 12*, 121–140.
Nie, X., & Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika, 108*(2), 299–319.
Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606.
Palmer, R., Roos, C., Vafaie, N., & Kober, H. (2023). The effect of ten versus twenty minutes of mindfulness meditation on state mindfulness and affect. *Scientific Reports, 13*(1), Article 20646.
Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., Cella, D., & PROMIS Cooperative Group. (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.
Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*(1), 160–164.
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.
Schumer, M. C., Lindsay, E. K., & Creswell, J. D. (2018). Brief mindfulness training for negative affectivity: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 86*(7), 569.
Schwartz, B., Giesemann, J., Delgadillo, J., Schaffrath, J., Hehlmann, M. I., Moggia, D., … Lutz, W. (2025). Comparing three neural networks to predict depression treatment outcomes in psychological therapies. *Behaviour Research and Therapy*, Article 104752.
Sparacio, A., Ijzerman, H., Ropovik, I., Giorgini, F., Spiessens, C., Uchino, B. N., … Jiga-Boy, G. M. (2024). Self-administered mindfulness interventions reduce stress in a large, randomized controlled multi-site study. *Nature Human Behaviour, 8*(9), 1716–1725.
Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer.
Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin, 95*(2), 334.
Steyerberg, E. W., & Harrell Jr, F. E. (2015). Prediction models need appropriate internal, internal-external, and external validation. *Journal of Clinical Epidemiology, 69*, 245.
Stieger, S., Schmid, I., Altenburger, P., & Lewetz, D. (2023). Short-, medium-, and long-term impact of watching humorous video clips on stress and well-being: An experience sampling method-based field experiment. *Computers in Human Behavior Reports, 10*, Article 100270.
Stone, C. J., Hansen, M. H., Kooperberg, C., & Truong, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling: 1994 wald memorial lecture. *Annals of Statistics, 25*(4), 1371–1470.
Sun, S., Simonsson, O., McGarvey, S., Torous, J., & Goldberg, S. B. (2024). Mobile phone interventions to improve health outcomes among patients with chronic diseases: An umbrella review and evidence synthesis from 34 meta-analyses. *The Lancet Digital Health, 6*, e857–e870.
Tanay, G., & Bernstein, A. (2013). State mindfulness scale (SMS): Development and initial validation. *Psychological Assessment, 25*(4), 1286.
Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*(2), 227–242.
van Agteren, J., Iasiello, M., Lo, L., Bartholomaeus, J., Kopsaftis, Z., Carey, M., & Kyrios, M. (2021). A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nature Human Behaviour, 5*(5), 631–652.
Vela, D., Sharp, A., Zhang, R., Nguyen, T., Hoang, A., & Pianykh, O. S. (2022). Temporal quality degradation in AI models. *Scientific Reports, 12*(1), Article 11654.
Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228–1242.
Wasil, A. R., Gillespie, S., Patel, R., Petre, A., Venturo-Conerly, K. E., Shingleton, R. M., … DeRubeis, R. J. (2020). Reassessing evidence-based content in popular smartphone apps for depression and anxiety: Developing and applying user-adjusted analyses. *Journal of Consulting and Clinical Psychology, 88*(11), 983.
Webb, C. A., Hirshberg, M. J., Davidson, R. J., & Goldberg, S. B. (2022). Personalized prediction of response to smartphone-delivered meditation training: Randomized controlled trial. *Journal of Medical Internet Research, 24*(11), Article e41566.
Webb, C. A., Ren, B., Hinze, V., Dalgleish, T., Ford, T. J., Greenberg, M. T., … Kuyken, W. (2026). Predicting adolescent response to school-based mindfulness: A secondary analysis of the MYRIAD trial. *JAMA Psychiatry*.
Webb, C. A., Swords, C. M., Lawrence, H. R., & Hilt, L. M. (2022). Which adolescents are well-suited to app-based mindfulness training? A randomized clinical trial and data-driven approach for personalized recommendations. *Journal of Consulting and Clinical Psychology, 90*(9), 655.
Webb, C. A., Swords, C. M., Murray, L., & Hilt, L. M. (2021). App-based mindfulness training for adolescent rumination: Predictors of immediate and cumulative benefit. *Mindfulness, 12*(10), 2498–2509.
Zantvoort, K., Nacke, B., Görlich, D., Hornstein, S., Jacobi, C., & Funk, B. (2024). Estimation of minimal data sets sizes for machine learning predictions in digital mental health interventions. *Npj Digital Medicine, 7*(1), 361.