Populated Pre-Analysis Plan

Advik Shreekumar Pierre-Luc Vautrey
University of California, Berkeley

January 8, 2026

This "populated pre-analysis plan" implements the analyses outlined in the pre-analysis plan, as recommended by Banerjee et al. (2020).

1 Treatments and Experimental Protocol

1.1 Recruitment and Sampling

We pre-registered the intent to recruit 2,500 participants into the randomization stage. Due to higher-than-expected recruitment costs, we halted at 2,384 participants to ensure our ability to pay all participants on incentivized tasks, even if every participant earned the maximum possible amount on all tasks.

1.2 Main Treatments and Auxiliary Manipulations

We successfully implemented our main randomization into 5 arms. We also successfully implemented our two auxiliary randomizations as specified (immediate meditation incentives for the treatment groups before the decision-making endline, and anxiety primers for all participants during the decision-making endline).

1.3 Experimental Protocol

We implemented all pre-specified surveys. After the final wave of participants completed the decision-making endline, we used our remaining funds to implement one additional survey, 3 to 4 months after randomization.

2 Empirical Analysis

2.1 Key Outcomes

- Meditation: see Table 1.
- Mental health: see Table 2 and Figure 1 for a debiased machine learning robustness check.

- Risk preferences: see Table 4.
- Productivity and Cognitive function: see Table 3 for ITT effects on the Stroop and proofreading tasks, and Table 4 for the interaction between the anxiety primer and meditation, and Figures 2 and 3 for debiased machine learning robustness checks.

2.2 Secondary Outcomes

- Information avoidance: see Table 6.
- Effects of cross-randomized anxiety primers on mood: see Table 5.
- Propensity to focus on the most relevant information: see Table 7; information avoidance appears separately above. We don't compute an index combining these outcomes because its interpretation depends too strongly on participants' assumed preferences.
- Time spent answering incentivized questions and tasks: we report these for the Stroop task (directly) and proofreading task (via productivity). For other incentivized tasks, our measurement of time spent includes time spent reading instructions due to how the survey was programmed; we omit these.
- Mindfulness scale (FFMQ-15): see Table 8. We don't perform mediation analyses, as treatment affected all FFMQ subscales. Performing a mediation analysis would require choosing a functional form to relate each subscale to our main outcomes, which we lack the information to credibly do. While planning, we anticipated effects on some but not all FFMQ subscales, which would have made this analysis more straightforward. For these reasons, we do not perform a mediation analysis.
- Beliefs about effects of meditation, perceived difficulty to meditate and willingness to pay for a license extension: see Tables 9, 10 and 11.

2.3 Regression Analysis

We present intent-to-treat analyses in the tables mentioned above. We present 2SLS analyses in Tables 12, 13 and 14, using three different first-stage outcomes: any app usage, days meditated, and minutes of meditation. We present debiased machine learning estimates in the tables mentioned above.

The relevant estimating equations are:

$$M = \beta_1 1\{App\} + \beta_2 1\{ShortIncentive\} + \beta_3\{LongIncentive\} + \alpha_s + \epsilon \tag{2.1}$$

$$Y_{post} = \beta_1 1\{App\} + \beta_2 1\{ShortIncentive\} + \beta_3\{LongIncentive\} + \alpha_s + \gamma Y_{pre} + \epsilon \tag{2.2}$$

$$M = \pi_1 1\{App\} + \pi_2 1\{ShortIncentive\} + \pi_3\{LongIncentive\} + \alpha_s^{(1)} + \gamma^{(1)} Y_{pre} + \nu \tag{2.3}$$

$$Y_{post} = \beta_1 \hat{M} + \alpha_s^{(2)} + \gamma^{(2)} Y_{pre} + \epsilon, \tag{2.4}$$

where the α are stratum fixed effects, M is a measurement of meditation activity (e.g., days meditated, or minutes meditated), Y is any outcome, and ϵ and ν are error terms.

2.4 Key Contrasts

We will investigate the following questions:

(i) Do short-term and long-term incentives aid habit formation? We will compare meditation outcomes in the short-term incentives group and in the long-term incentives group to meditation outcomes in the pure incentives group.

- See Table 1.

(ii) Does mindfulness meditation reduce anxiety and other mental health issues? We will compare GAD-7 scores, PSS scores and PHQ-8 scores among participants who meditate more and less, using treatment status as an instrument for meditation activity.

- See Table 2 and Tables 12, 13 and 14.

(iii) Does meditation improve the amount of attention and the ability to allocate attention correctly? We will compare scores on the Stroop task and the Proofreading task among participants who meditate more and less. We will also compare the extent to which being assigned to an anxiety primer alters performance on productivity tasks and decision-making on the Callen et al. 2014 task between participants who meditate more and less.

- See Table 3 and Table 4.

3 Tables

Table 1: App Usage and Habit Formation

	(1)	(2)	(3)	(4)	(5)
	Time from Randomization				
	Days 1-16	Days 17-28	Days 1-28	Day 29+	Cumulative
A. Minutes Per Day					
App Access	5.29***	4.08***	4.77***	1.04**	2.20***
(S.E.)	(0.44)	(0.38)	(0.37)	(0.41)	(0.36)
Short Incentive	1.80***	0.31	1.16**	−0.13	0.27
(S.E.)	(0.57)	(0.54)	(0.51)	(0.51)	(0.46)
Long Incentive	3.28***	0.35	2.02***	−0.73	0.13
(S.E.)	(0.60)	(0.52)	(0.52)	(0.45)	(0.43)
Waitlist Mean	0.099	0.034	0.134	91.788	91.922
N	2384	2384	2384	2384	2384
B. Days per Week					
App Access	2.65***	1.29***	1.96***	0.19**	0.74***
(S.E.)	(0.11)	(0.08)	(0.09)	(0.08)	(0.07)
Short Incentive	0.66***	0.01	0.33***	−0.05	0.07
(S.E.)	(0.15)	(0.11)	(0.12)	(0.10)	(0.09)
Long Incentive	1.36***	0.14	0.75***	−0.12	0.15*
(S.E.)	(0.16)	(0.11)	(0.12)	(0.09)	(0.09)
Waitlist Mean	0.015	0.006	0.021	4.497	4.518
N	2384	2384	2384	2384	2384

Notes: This table presents the average treatment effects of app access and usage incentives on app usage during various time windows. We calculate usage based on administrative data associated with each participant's unique voucher code. Panel A presents effects on minutes meditated per day, and Panel B does the same for days meditated per week. The estimating equation is Equation (2.1), which includes stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Effects on Mental Health

	Time from Randomization				
	Four Days	Seven Days	Eleven Days	Two Weeks	Four Weeks
	(1)	(2)	(3)	(4)	(5)
App Access	-0.089^{**}	-0.157^{***}	-0.226^{***}	-0.381^{***}	-0.456^{***}
(S.E.)	(0.040)	(0.042)	(0.042)	(0.036)	(0.039)
Short Incentive	-0.015	-0.010	-0.061	-0.081^{*}	-0.047
(S.E.)	(0.048)	(0.049)	(0.050)	(0.043)	(0.045)
Long Incentive	-0.110^{**}	-0.064	-0.078	-0.070^{*}	-0.046
(S.E.)	(0.046)	(0.050)	(0.051)	(0.042)	(0.045)
Waitlist Mean	0.017	0.017	0.034	0.035	0.056
N	2305	2145	2191	2330	2311
Index Components:					
Anxiety	GAD-2	GAD-2	GAD-2	GAD-7	GAD-7
Depression				PHQ-8	PHQ-8
Stress				PSS-10	PSS-10

Notes: This table presents average treatment effects of app access and usage incentives on reported symptoms of mental distress over time. We measure symptoms of anxiety using the two- and seven-item Generalized Anxiety Disorder scales (GAD-2 and GAD-7, respectively); symptoms of depression using the eight-item Patient Health Questionnaire (PHQ-8); and stress using the ten-item Perceived Stress Scale (PSS-10). The outcome at each timepoint is a standardized index that combines the mental health scales measured at that time. We first standardize each scale in each time period by subtracting the Pure Waitlist mean and dividing by the Pure Waitlist standard deviation. The index is the average of these standardized scales. Lower scores indicate lower reported levels of distress. The estimating equation is Equation (2.2), which includes stratum fixed effects and the baseline mental health index. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table 3: Effects on Attention and Productivity

	(1) Stroop Errors Made	(2) Stroop Time Taken	(3) Proofreading Errors Found	(4) Proofreading Productivity
App Access	-0.132	-0.029	0.240^{*}	0.002
(S.E.)	(0.169)	(0.394)	(0.135)	(0.003)
Short Incentive	0.196	0.046	0.001	-0.004
(S.E.)	(0.198)	(0.585)	(0.152)	(0.003)
Long Incentive	0.068	0.170	-0.011	0.001
(S.E.)	(0.181)	(0.477)	(0.151)	(0.003)
Waitlist Mean	1.439	67.817	14.756	0.135
N	2256	2254	2257	2257

Notes: This table presents average treatment effects of app access on performance in two incentivized tasks: a Stroop test and a proofreading task. The estimating equation is Equation (2.2), which includes stratum fixed effects and performance on the given task in the baseline survey. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table 4: Interference of Anxiety on Decision Making

	(1) Proofreading Errors Found	(2) Proofreading Productivity	(3) Indifference Point (Certainty)	(4) Certainty Premium
App Access × Neutral Memory	0.340**	0.000	1.397	−0.031*
(S.E.)	(0.142)	(0.003)	(1.228)	(0.018)
Stressful Memory	−0.060	−0.001	0.704	−0.038**
(S.E.)	(0.168)	(0.003)	(1.310)	(0.020)
App Access × Stressful Memory	0.134	0.002	0.144	0.029
(S.E.)	(0.152)	(0.003)	(1.212)	(0.018)
Waitlist Mean	14.830	0.139	58.030	0.163
N	2257	2257	2247	2247

Notes: We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Subjective Mood after Stressful Tasks

	(1)	(2)
	Mood After . . .	
	First Task	Second Task
App Access × Neutral Task	−0.540***	−0.422***
(S.E.)	(0.060)	(0.057)
Stressful Task	1.046***	0.306***
(S.E.)	(0.068)	(0.065)
App Access × Stressful Task	−0.098	−0.207***
(S.E.)	(0.065)	(0.060)
Waitlist Mean	1.966	2.032
N	2257	2257

Notes: This table presents average treatment effects of the Stressful tasks on self-reported mood. The stressful task involves recounting either an unresolved worry (Stressful) or a daily routine (Neutral). The second task involves describing how one would respond to a large medical bill (Stressful) or a small one (Neutral). After each task, we elicit participants' mood on a 6-point scale from "very calm, very relaxed" (1) to "very upset, very stressed" (6). All regressions include stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Avoidance of Potentially Distressing Information

	(1) Index	(2) Life Expectancy	(3) Dementia	(4) Job Loss	(5) Retirement Finances
App Access	0.006	0.013	−0.009	0.010	0.009
(S.E.)	(0.019)	(0.027)	(0.027)	(0.028)	(0.029)
Short Incentive	−0.014	0.012	−0.013	−0.054	−0.002
(S.E.)	(0.023)	(0.032)	(0.031)	(0.033)	(0.033)
Long Incentive	−0.007	0.005	−0.016	−0.016	−0.002
(S.E.)	(0.023)	(0.032)	(0.031)	(0.033)	(0.033)
Waitlist Mean	0.454	0.319	0.328	0.595	0.576
N	2257	2257	2257	2257	2257

Notes: This table presents the average treatment effects on avoidance of potentially distressing information. We offer participants optional informational links, which are ultimately delivered at the end of the decision-making survey. They can choose to receive up to four links: (i) a life expectancy calculator; (ii) risk factors for developing dementia; (iii) the risk of one's job of being replaced by automation; and (iv) a calculator of financial risk in retirement. The index is the proportion of links the participant refuses. All regressions include stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Focus on Relevant Information

	(1) Demand Relevant Info	(2) Correct Choice Given Info	(3) Reject Low Prob. Loss
App Access	−0.034	−0.028	0.021
(S.E.)	(0.027)	(0.027)	(0.027)
Short Incentive	0.001	0.045	−0.003
(S.E.)	(0.032)	(0.031)	(0.032)
Long Incentive	−0.012	−0.001	0.046
(S.E.)	(0.032)	(0.031)	(0.032)
Waitlist Mean	0.706	0.736	0.319
N	2257	2158	2257

Notes: This table presents the average treatment effects on take-up of relevant information in incentivized tasks. Columns 1 and 2 refer to choices in a simple gamble: participants are endowed with a $0 or $1 bonus on top of their payment for completing the survey with equal probabilit ($ 10). They can then choose to take a gamble that will either add or subtract a further $1 from their total payment. The gamble involves three virtual coin tosses. If at least two tosses are heads, the participant wins and additional dollar; if at least two are tails, the participant loses an additional dollar. Before taking the gamble, participants can choose to know the result of the first coin toss ("relevant"), whether their bonus is $0 or $1 ("irrelevant"), or the age of the oldest tree in the world ("irrelevant"). Participants receive all pieces of information with 95% probability, and receive only their selected piece of information with 5% probability. The outcome in column 1 is an indicator for whether participants selected information about the coin toss; and the outcome in column 2 is an indicator for whether participants elect to take (or avoid) the bet when the first coin turned heads (versus tails). The latter choice is "correct" in the sense of maximizing expected utiltiy for a risk-neutral agent. Then, participants are presented with the choice to gamble their base payment ($ 10) in a lottery that gives them $1 with probability 0.99 but takes away their base payment with probability 0.01. Column 3 presents the proportion of participants who reject this gamble. All regressions include stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Five-Factor Mindfulness Questionnaire

	(1) Total	(2) Observing	(3) Describing	(4) Acting with Awareness	(5) Non-judgment	(6) Non-reactivity
App Access	0.194***	0.228***	0.150***	0.184***	0.240***	0.172***
(S.E.)	(0.022)	(0.036)	(0.034)	(0.033)	(0.039)	(0.036)
Short Incentive	0.022	−0.022	0.016	0.043	0.041	0.033
(S.E.)	(0.026)	(0.042)	(0.039)	(0.039)	(0.043)	(0.042)
Long Incentive	−0.005	−0.018	0.016	0.002	−0.011	−0.009
(S.E.)	(0.026)	(0.043)	(0.039)	(0.038)	(0.045)	(0.042)
Waitlist Mean	2.232	2.164	2.484	1.961	2.472	2.080
N	2330	2330	2330	2330	2330	2330

Notes: This table presents average treatment effects of app access and usage incentives on mindfulness, as measured by the Five Facet Mindfulness Questionnaire (FFMQ-15). The FFMQ-15 contains fifteen items, each of which is scored on an integer scale of 1 (lower mindfulness) to 5 (higher mindfulness). The fifteen items are commonly grouped into 5 subscales of 3 items each, which pertain to different aspects of mindfulness. Column 1 presents the average score across all 15 items. Columns 2 through 6 present effects on the subscales. All regressions include stratum fixed effects and control for a baseline value of the outcome. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Subjective Beliefs about Treatment Effect on Anxiety

	(1) 4 Days	(2) 7 Days	(3) 11 Days	(4) 4 Weeks
App Access	−0.060	−0.069	0.054	0.284***
(S.E.)	(0.073)	(0.095)	(0.099)	(0.102)
Short Incentive	0.058	−0.005	−0.049	−0.143
(S.E.)	(0.090)	(0.109)	(0.115)	(0.123)
Long Incentive	0.071	0.194*	0.101	0.124
(S.E.)	(0.087)	(0.111)	(0.119)	(0.118)
Waitlist Mean	−3.589	−3.551	−3.565	−3.594
N	2289	2131	2185	2309

Notes: This table presents average treatment effects of app access and usage incentives on participants' predictions about the effect of mindfulness meditation on mental health. To measure subjective treatment effects, we ask participants to consider a hypothetical scenario. We instruct them to consider 10 other randomly selected participants who report anxiety symptoms at the beginning of the study. Then, we ask them to predict how many of these 10 would report anxiety in 3 weeks if they did not receive a Headspace license ("control"), as well as if they did receive the license and used it for 5 or more days per week ("treatment"). Finally, we calculate participants' subjective treatment effect as the treatment-minus control improvement, which is an integer ranging from -10 to 10. These questions appear at baseline, as well as 4, 7, 11, and 30 days post-randomization. In all post-baseline elicitations, we remind participants of their previous responses and present an opportunity to update them. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Subjective Willingness to Pay for Voucher Extension

	(1) 4 Days	(2) 7 Days	(3) 11 Days	(4) 4 Weeks
App Access	2.945***	1.760	1.633	−0.013
(S.E.)	(1.063)	(1.165)	(1.200)	(1.368)
Short Incentive	0.579	1.833	1.166	0.326
(S.E.)	(1.306)	(1.389)	(1.481)	(1.640)
Long Incentive	0.229	2.300*	1.633	0.364
(S.E.)	(1.233)	(1.351)	(1.426)	(1.660)
Pure Waitlist Mean	50.097	50.372	50.329	49.615
N	2260	2140	2189	2310

Notes: This table presents average treatment effects of app access and usage incentives on willingness to pay for a 90-day extension of the Headspace license. The cash value of the license was approximately $39 at the time of the experiment. We elicit willingness to pay with a probabilistic Becker–DeGroot–Marschak mechanism, allowing participants to indicate a valuation for the extension of between $0 and $100 using a sliding scale. This question appears on surveys at baseline, as well as post-randomization at 4, 7, 11, and 30 days. We implement the mechanism with 1% probability per participant, selecting one of their responses uniformly at random after excluding missing values (say, if they skipped a survey with the WTP elicitation). All regressions include stratum fixed effects and control for a baseline value of the outcome. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Ease of Meditating (Treatment Group Only)

	(1) 4 Days	(2) 7 Days	(3) 11 Days	(4) 3-4 Months
A. Finding Time and Space				
Short Incentive	0.498***	0.413**	0.659***	0.438**
(S.E.)	(0.159)	(0.176)	(0.177)	(0.185)
Long Incentive	0.450***	0.268	0.106	0.068
(S.E.)	(0.159)	(0.172)	(0.173)	(0.186)
App Access Only Mean	4.696	4.919	4.905	5.012
Sample Size	1098	825	816	956
B. Focusing for 10 Minutes				
Short Incentive	0.044	0.088***	0.067**	0.053*
(S.E.)	(0.032)	(0.032)	(0.032)	(0.030)
Long Incentive	0.062**	0.071**	0.066**	0.044
(S.E.)	(0.031)	(0.033)	(0.032)	(0.030)
App Access Only Mean	0.603	0.618	0.634	0.672
Sample Size	1352	1237	1264	1367

Notes: This table presents intent-to-treat effects of offering usage incentives on self-reported ease of meditating. We collect responses only from the 1,429 participants in the treatment group and use the estimating equation $Y_i = \delta_{\text{stratum}} + \beta_1 \text{ShortIncentive}_i + \beta_2 \text{LongIncentive}_i + \epsilon_i$ at each time point. In Panel A, the outcome comes from answers to the question "In your experience so far, how easy or difficult is it to find a good time and space to meditate?", measured on a scale from 0 ("very difficult") to 10 ("very easy"). We treat this as an integer between 0 and 10. In Panel B, the outcome is the share of participants who responded that it was "very easy" or "somewhat easy" to focus on meditating for 10 minutes without quitting, given the right time and space. The other options were "very difficult" "somewhat difficult", and "I don't know—I have not been meditating". We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: LATE of Any App Usage on Key Outcomes

	(1)	(2)	(3)	(4)
	Two Weeks	**Four Weeks**		**Three-Four Months**
	Mental Health Index	Mental Health Index	Proofreading	Mental Health Index
A. App Access (Pooled)				
Any App Usage	-0.492***	-0.542***	1.366**	-0.472***
(S.E.)	(0.029)	(0.031)	(0.588)	(0.074)
[95% AR Interval]	[-0.572, -0.412]	[-0.631, -0.454]	[-0.239, 2.975]	[-0.681, -0.270]
B. App Access Only				
Any App Usage	-0.468***	-0.534***	1.545*	-0.472***
(S.E.)	(0.044)	(0.046)	(0.807)	(0.111)
[95% AR Interval]	[-0.552, -0.385]	[-0.623, -0.447]	[-0.143, 3.236]	[-0.700, -0.254]
C. Short Incentives				
Any App Usage	-0.510***	-0.548***	1.271*	-0.450***
(S.E.)	(0.039)	(0.042)	(0.759)	(0.094)
[95% AR Interval]	[-0.585, -0.435]	[-0.629, -0.467]	[-0.278, 2.821]	[-0.646, -0.261]
D. Long Incentives				
Any App Usage	-0.491***	-0.542***	1.283*	-0.496***
(S.E.)	(0.038)	(0.041)	(0.739)	(0.093)
[95% AR Interval]	[-0.565, -0.418]	[-0.621, -0.462]	[-0.239, 2.809]	[-0.691, -0.308]
E. Supplementary Information				
Sargan-Hansen Overidentification Test				
J-statistic	0.73	0.07	0.10	0.16
p value	0.694	0.967	0.951	0.925
First-Stage F				
Pooled	2379	2800	2470	228
App Access Only	4075	5208	4230	233
Short Incentives	9039	9870	9553	323
Long Incentives	10227	10921	10394	321
Sample Size				
Pooled	2330	2311	2257	2004
App Access Only	1408	1398	1373	1209
Short Incentives	1400	1389	1366	1209
Long Incentives	1408	1396	1368	1202

Notes: This table presents estimates of the Local Average Treatment Effect (LATE) of using the app at all on key outcomes. The endogenous variable is an indicator for completing at least one meditation session on the app before taking a given followup survey. The estimating equations are 2.3 for the first stage and 2.4 for the second stage. Panel A presents estimates that instrument for app usage with all three treatment arms. Panels B, C, and D focus on the LATE for each treatment arm separately. Each panel provides a point-estimate, a heteroskedasticity-robust standard error, and a pre-specified 95% confidence interval from the Anderson-Rubin procedure for robustness to weak instruments. Panel E presents additional information useful for interpreting two-stage least squares regressions. First, it provides a Sargan-Hansen overidentification test for Panel A, which roughly corresponds to testing whether every instrument implies the same LATE. Observing a large J-statistic (or small p-value) implies that the AR confidence interval in Panel A is likely unreliable. Panel E also provides the first-stage F statistic and number of observations for the regressions in panels A through D.

Table 13: LATE of Days Meditated on Key Outcomes

	(1) **Two Weeks**	(2) **Four Weeks**	(3)	(4) **Three-Four Months**
	Mental Health Index	Mental Health Index	Proofreading	Mental Health Index
A. App Access (Pooled)				
Days Per Week	-0.426***	-0.332***	1.110**	-0.125***
(S.E.)	(0.026)	(0.020)	(0.494)	(0.020)
[95% AR Interval]	[-0.438, -0.415]	[-0.361, -0.306]	[-0.209, 2.432]	[-0.184, -0.071]
B. App Access Only				
Days Per Week	-0.493***	-0.377***	1.559*	-0.124***
(S.E.)	(0.047)	(0.034)	(0.817)	(0.030)
[95% AR Interval]	[-0.583, -0.405]	[-0.442, -0.314]	[-0.145, 3.275]	[-0.185, -0.066]
C. Short Incentives				
Days Per Week	-0.474***	-0.357***	1.136*	-0.128***
(S.E.)	(0.037)	(0.028)	(0.677)	(0.027)
[95% AR Interval]	[-0.544, -0.404]	[-0.411, -0.304]	[-0.249, 2.521]	[-0.184, -0.075]
D. Long Incentives				
Days Per Week	-0.385***	-0.303***	0.964*	-0.124***
(S.E.)	(0.030)	(0.023)	(0.554)	(0.024)
[95% AR Interval]	[-0.443, -0.328]	[-0.348, -0.258]	[-0.180, 2.111]	[-0.173, -0.077]
E. Supplementary Information				
Sargan-Hansen Overidentification Test				
J-statistic	7.66	6.01	0.51	0.03
p value	0.022	0.050	0.775	0.985
First-Stage F				
Pooled	729	542	720	108
App Access Only	1216	1003	1196	184
Short Incentives	2395	1715	2282	243
Long Incentives	2635	2064	2712	335
Sample Size				
Pooled	2330	2311	2257	2004
App Access Only	1408	1398	1373	1209
Short Incentives	1400	1389	1366	1209
Long Incentives	1408	1396	1368	1202

Notes: This table presents estimates of the Local Average Treatment Effect (LATE) of days using the app on key outcomes. The endogenous variable is the number of days on which a participant completed a meditation session on the app, counting only sessions that took place before the survey in question. The estimating equations are 2.3 for the first stage and 2.4 for the second stage. Panel A presents estimates that instrument for app usage with all three treatment arms. Panels B, C, and D focus on the LATE for each treatment arm separately. Each panel provides a point-estimate, a heteroskedasticity-robust standard error, and a pre-specified 95% confidence interval from the Anderson-Rubin procedure for robustness to weak instruments. Panel E presents additional information useful for interpreting two-stage least squares regressions. First, it provides a Sargan-Hansen overidentification test for Panel A, which roughly corresponds to testing whether every instrument implies the same LATE. Observing a large J-statistic (or small p-value) implies that the AR confidence interval in Panel A is likely unreliable. Panel E also provides the first-stage F statistic and number of observations for the regressions in panels A through D.

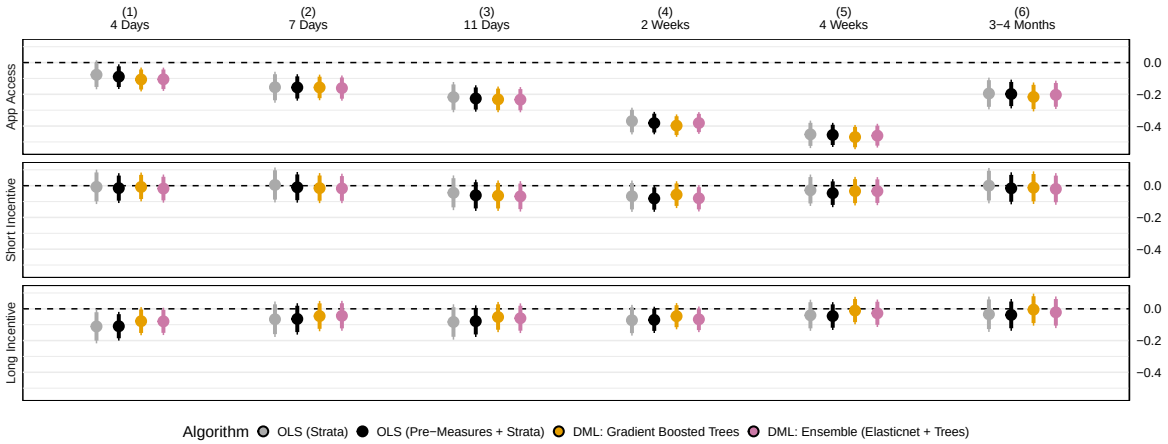
Table 14: LATE of 10 Minutes of Meditation on Key Outcomes

	(1) **Two Weeks**	(2) **Four Weeks**	(3)	(4) **Three-Four Months**
	Mental Health Index	Mental Health Index	Proofreading	Mental Health Index
A. App Access (Pooled)				
10 Minutes Per Day	-0.006***	-0.008***	0.016**	-0.008***
(S.E.)	(0.000)	(0.001)	(0.007)	(0.001)
[95% AR Interval]	[-0.006, -0.006]	[-0.009, -0.007]	[-0.003, 0.035]	[-0.013, -0.005]
B. App Access Only				
10 Minutes Per Day	-0.007***	-0.010***	0.024*	-0.008***
(S.E.)	(0.001)	(0.001)	(0.013)	(0.002)
[95% AR Interval]	[-0.009, -0.006]	[-0.012, -0.008]	[-0.002, 0.050]	[-0.012, -0.004]
C. Short Incentives				
10 Minutes Per Day	-0.007***	-0.009***	0.016*	-0.008***
(S.E.)	(0.001)	(0.001)	(0.010)	(0.002)
[95% AR Interval]	[-0.008, -0.006]	[-0.010, -0.007]	[-0.004, 0.036]	[-0.012, -0.004]
D. Long Incentives				
10 Minutes Per Day	-0.005***	-0.008***	0.014*	-0.009***
(S.E.)	(0.000)	(0.001)	(0.008)	(0.002)
[95% AR Interval]	[-0.006, -0.005]	[-0.009, -0.006]	[-0.003, 0.031]	[-0.013, -0.005]
E. Supplementary Information				
Sargan-Hansen Overidentification Test				
J-statistic	7.59	4.94	0.61	0.27
p value	0.022	0.085	0.736	0.873
First-Stage F				
Pooled	214	179	210	37
App Access Only	285	323	281	75
Short Incentives	778	579	751	95
Long Incentives	860	677	868	110
Sample Size				
Pooled	2330	2311	2257	2004
App Access Only	1408	1398	1373	1209
Short Incentives	1400	1389	1366	1209
Long Incentives	1408	1396	1368	1202

Notes: This table presents estimates of the Local Average Treatment Effect (LATE) of minutes using the app on key outcomes. The endogenous variable is the number of minutes meditated using the app, counting only sessions that took place before the survey in question. We divide the number of minutes by 10, which is the length of the typical introductory meditation session on the app. The estimating equations are **??** for the first stage and 2.4 for the second stage. Panel A presents estimates that instrument for app usage with all three treatment arms. Panels B, C, and D focus on the LATE for each treatment arm separately. Each panel provides a point-estimate, a heteroskedasticity-robust standard error, and a pre-specified 95% confidence interval from the Anderson-Rubin procedure for robustness to weak instruments. Panel E presents additional information useful for interpreting two-stage least squares regressions. First, it provides a Sargan-Hansen overidentification test for Panel A, which roughly corresponds to testing whether every instrument implies the same LATE. Observing a large J-statistic (or small p-value) implies that the AR confidence interval in Panel A is likely unreliable. Panel E also provides the first-stage F statistic and number of observations for the regressions in panels A through D.

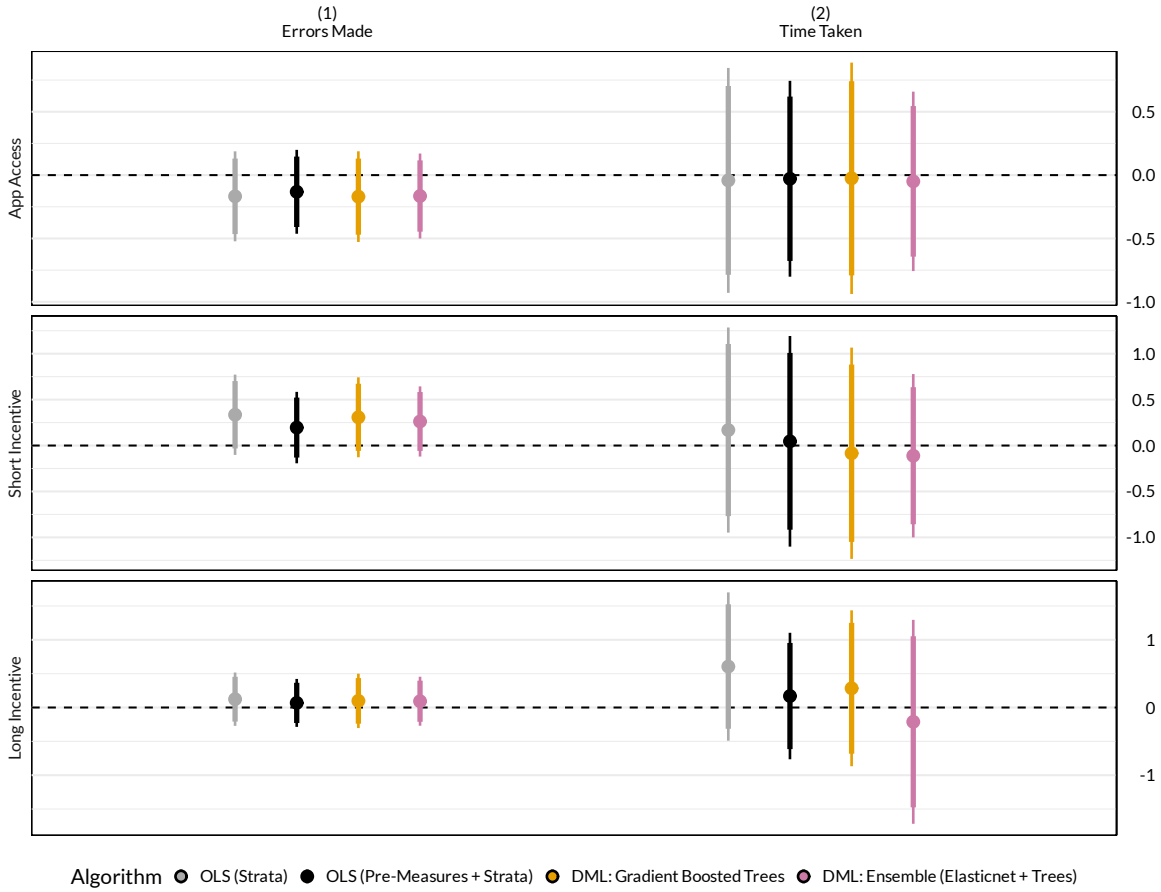
4 Figures

Figure 1: Effects of App Access and Usage Incentives on Mental Health, Adjusting for Covariates



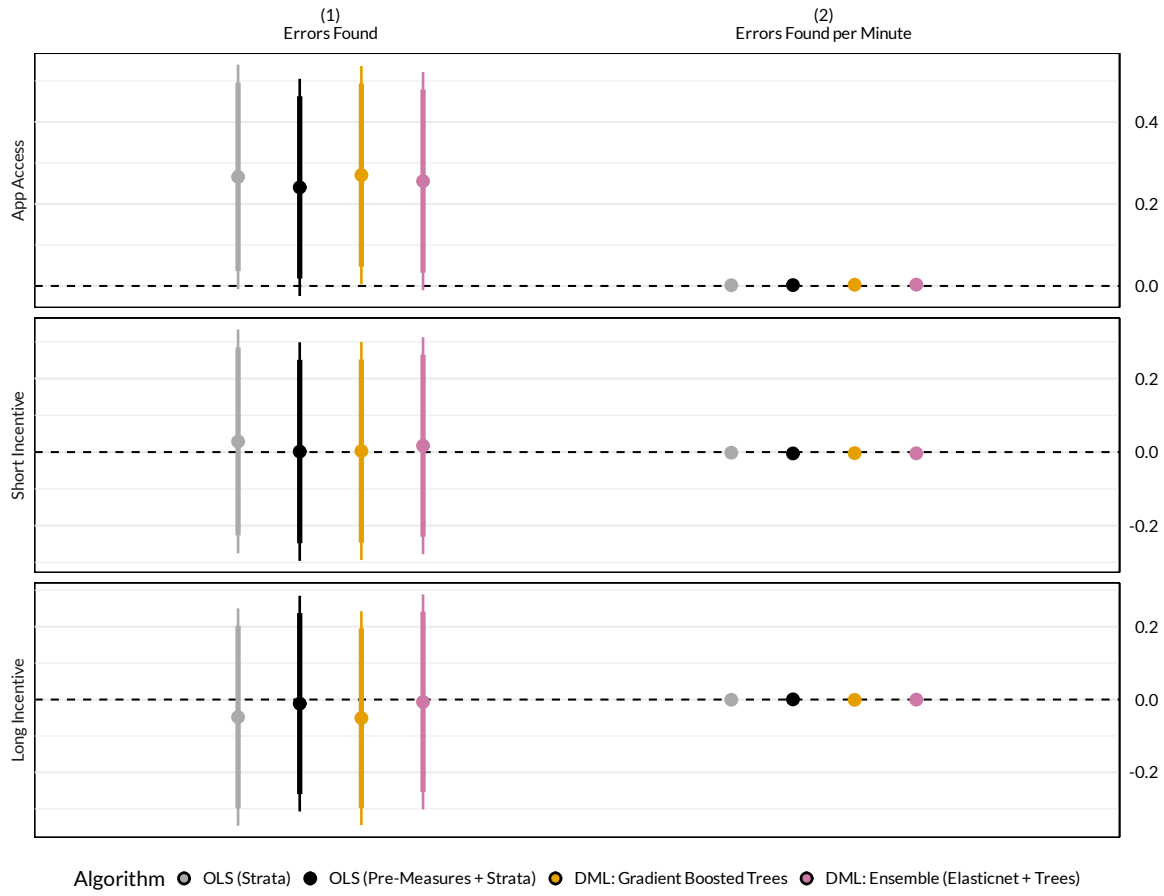
Notes: This figure presents average treatment effects of app access and usage incentives on reported symptoms of metal distress over time, adjusting for covariates using debiased machine learning (Chernozhukov et al., 2018). The figure's layout mirrors that of Table 2. In each column, the black point corresponds to the point estimate reported in Table 2, the thick bar corresponds to a 90% confidence interval, and the thin bar a 95% confidence interval. This point estimate comes from Equation 2.2, which includes stratum fixed effects and a baseline measure of the outcome. The remaining points present the debiased machine learning estimates of the same parameters, using 5-fold cross validation and one of two algorithms: gradient boosted trees, and ensemble that uses gradient-boosted trees to predict the residuals from an elasticnet regression.

Figure 2: Effects of App Access and Usage Incentives on Stroop Task, Adjusting for Covariates



Notes: This figure presents average treatment effects of app access and usage incentives on Stroop task performance, adjusting for covariates using debiased machine learning (Chernozhukov et al., 2018). The figure's layout mirrors that of Table 3, columns 1 and 2. In each column, the black point corresponds to the point estimate reported in Table 3, the thick bar corresponds to a 90% confidence interval, and the thin bar a 95% confidence interval. This point estimate comes from Equation 2.2, which includes stratum fixed effects and a baseline measure of the outcome. The remaining points present the debiased machine learning estimates of the same parameters, using 5-fold cross validation and one of two algorithms: gradient boosted trees, and ensemble that uses gradient-boosted trees to predict the residuals from an elasticnet regression.

Figure 3: Effects of App Access and Usage Incentives on Proofreading Task, Adjusting for Covariates



Notes: This figure presents average treatment effects of app access and usage incentives on proofreading task performance, adjusting for covariates using debiased machine learning (Chernozhukov et al., 2018). The figure's layout mirrors that of Table 3, columns 1 and 2. In each column, the black point corresponds to the point estimate reported in Table 3, the thick bar corresponds to a 90% confidence interval, and the thin bar a 95% confidence interval. This point estimate comes from Equation 2.2, which includes stratum fixed effects and a baseline measure of the outcome. The remaining points present the debiased machine learning estimates of the same parameters, using 5-fold cross validation and one of two algorithms: gradient boosted trees, and ensemble that uses gradient-boosted trees to predict the residuals from an elasticnet regression.