BRIEF REPORT

An exploratory application of machine learning methods to optimize prediction of responsiveness to digital interventions for eating disorder symptoms

Jake Linardon PhD[1] 🔾 | Matthew Fuller-Tyszkiewicz PhD[1] | Adrian Shatte PhD[2] | Christopher J. Greenwood PhD[1,3]

[1]Centre for Social and Early Emotional Development and School of Psychology, Deakin University, Geelong, Victoria

[2]Federation University, School of Engineering, Information Technology & Physical Sciences, Melbourne, Victoria, Australia

[3]Centre for Adolescent Health, Murdoch Children's Research Institute, Parkville, Victoria, Australia

Correspondence
Jake Linardon, Centre for Social and Early Emotional Development and School of Psychology, Deakin University, Geelong, Victoria.
Email: jake.linardon@deakin.edu.au

Abstract

Objective: Digital interventions show promise to address eating disorder (ED) symptoms. However, response rates are variable, and the ability to predict responsiveness to digital interventions has been poor. We tested whether machine learning (ML) techniques can enhance outcome predictions from digital interventions for ED symptoms.

Method: Data were aggregated from three RCTs ($n = 826$) of self-guided digital interventions for EDs. Predictive models were developed for four key outcomes: uptake, adherence, drop-out, and symptom-level change. Seven ML techniques for classification were tested and compared against the generalized linear model (GLM).

Results: The seven ML methods used to predict outcomes from 36 baseline variables were poor for the three engagement outcomes (AUCs = 0.48–0.52), but adequate for symptom-level change (R^2 = .15–.40). ML did not offer an added benefit to the GLM. Incorporating intervention usage pattern data improved ML prediction accuracy for drop-out (AUC = 0.75–0.93) and adherence (AUC = 0.92–0.99). Age, motivation, symptom severity, and anxiety emerged as influential outcome predictors.

Conclusion: A limited set of routinely measured baseline variables was not sufficient to detect a performance benefit of ML over traditional approaches. The benefits of ML may emerge when numerous usage pattern variables are modeled, although this validation in larger datasets before stronger conclusions can be made.

KEYWORDS
adherence, digital, eating disorders, e-health, engagement, intervention, machine learning, prediction, randomized controlled trial, uptake

1 | INTRODUCTION

The enthusiasm for digital interventions to address eating disorders (ED) is mounting. Digital interventions can alleviate existing help-seeking barriers (Weissman & Rosselli, 2017). However, response rates to digital interventions are variable, with symptom remission occurring in only 20%–50% of cases (Fitzsimmons-Craft et al., 2020). Using commonly collected baseline and process data to develop

accurate predictive models capable of identifying likelihood of success is necessary for improving patient outcomes (Chekroud et al., 2021).

Machine learning (ML) may help with developing accurate prediction models. ML involves data-driven techniques that enable computer algorithms to identify and iteratively refine the optimal parameters to fit complex variable patterns (Jordan & Mitchell, 2015). ML is suited to situations where there are a large number of predictors to model and the best combination of predictors is uncertain a priori. ML enables prediction models to be cross-validated by comparing precision and accuracy across training and test subsets of a single dataset, or with external validation datasets for greater out-of-sample generalizability (Shalev-Shwartz & Ben-David, 2014).

The value of ML may depend on several factors. When assumptions and sample size requirements are satisfied, the number of predictor variables is small, and nonlinear effects are weak, traditional techniques produce prediction models as accurate as ML (Christodoulou et al., 2019). However, when the number of predictors is large and there are unanticipated nonlinear interactions, the incremental benefits of ML become apparent (Pearson et al., 2019).

ML techniques have been applied in the context of digital interventions. Accurate ML prediction models for program engagement have been observed in online mindfulness programs for wellbeing (Lekkas et al., 2021), internet-based cognitive therapy for depression (Chien et al., 2020), and web-based behavioral programs for insomnia (Bremer et al., 2020). In digital interventions for body dysmorphic disorder, ML has resulted in superior prediction models to traditional techniques when predicting symptom-level change from baseline and process variables (Flygare et al., 2020).

We applied ML to predict outcomes from baseline data collected in RCTs of digital interventions for ED symptoms. Our exploratory aims were to: (1) generate accurate ML-based predictive models using mostly baseline data; (2) determine whether ML enhances prediction over traditional techniques; and (3) explore influential outcome predictors that will inform future confirmatory work.

2 | METHOD

2.1 | Study design

We aggregated data from 848 participants enrolled in three RCTs of self-guided digital interventions for ED symptoms. Details of these RCTs have been published elsewhere (Linardon, Messer, Shatte, Greenwood, et al., 2021; Linardon, Messer, Shatte, Skvarc, et al., 2021; Linardon, Shatte, Rosato, & Fuller-Tyszkiewicz, 2020). There were five participant groups with pre- and 4-week post digital intervention exposure data. Ethics approval and informed consent were obtained.

2.2 | Study population and recruitment

The population and recruitment method were nearly identical across the three trials. Participants were recruited via advertisements distributed throughout the first author's psychoeducational ED platform. This platform consists of an open-access website and social media accounts that offer educational material related to EDs (for more detail about this platform, see Linardon, Rosato, & Messer, 2020).

There was one difference in study inclusion criteria between trials. Participants needed to report the presence of at least one objective binge eating episode in one trial, whereas this criterion was not employed in the other two. All participants were required to be aged 18 years or over.

The final sample used in this study included 826 individuals who provided data on all engagement outcomes and baseline predictors. Three-hundred and fifty-nine participants provided pre- and posttest data on symptom-level change.

2.3 | Digital interventions

There were two digital intervention programs delivered across the trials, *Break Binge Eating* (two groups; 1 trial) and *Breaking the Diet Cycle* (three groups; 2 trials). Both programs were structurally the same; that is, they both contained four learning modules/sessions, were self-guided and based on CBT principles, offered similar homework exercises and functionality (symptom monitoring, quizzes, progress monitoring, reflection tasks, etc.), and only slightly differed in program length. There were two key differences between the programs. The first was that *Break Binge Eating* was delivered solely through a smartphone app, while *Breaking the Diet Cycle* was delivered through both a web platform and an app. The second was that *Break Binge Eating* targeted multiple maintaining mechanisms (restriction, body image, and mood dysregulation) while *Breaking the Diet Cycle* targeted dietary restriction only. Given the similarities between the two programs, we deemed it appropriate to amalgamate them in the analyses. Please refer to Table S1 for a description of these programs.

2.4 | Measures

2.4.1 | Outcomes

Four outcomes were selected for analyses: uptake; adherence, dropout, and adherence (see Supporting Information Materials S1 for their operationalization).

2.4.2 | Baseline predictors

Thirty-six predictors were used, divided into three clusters: demographic, psychiatric and treatment; and symptom severity. For the list of baseline variables, see the Supporting Information Materials S1.

2.4.3 | Usage behavior predictors

In one trial ($n = 340$), 110 usage pattern variables were available for analysis (Table S2). These were modeled in sensitivity analyses.

2.4.4 | Data analysis

Predictive models were generated using eight classification approaches: (1) traditional regression, (2) elastic-net penalized regression, (3) support vector machine with linear kernel, (4) support vector machine with polynomial kernel, (5) support vector machine with radial basis function kernel, (6) k-Nearest Neighbor, (7) Classification and Regression Tree (CART), and (8) random forest. Predictive models were implemented using the *caret* package (Kuhn, 2021). Predictors were all centered and standardized.

An iterative process was implemented to examine predictive performance and reduce data biases, whereby the full dataset was split into 100 different training (67%) and testing (33%) datasets. Each model was generated in the training data and then validated in the testing data. Models were trained using fivefold cross-validation to select the optimal model. To quantify predictive performance from each validation set, accuracy (total correct predictions divided by all predictions), area under the curve (AUC), $F1$ score, true negatives, false positives, false negatives, and true positives were computed for binary outcomes (engagement indices), and R^2, Root-Mean-Square Error, and Mean Absolute Error were computed from the continuous outcome. Each predictive performance index was averaged across all iterations. Variable importance was extracted by running the most accurate model on the full dataset.

Sensitivity analyses were conducted using 10 iterations of training–testing splits. For binary outcomes (engagement indices), models were repeated using: (1) an artificially oversampled dataset (synthetic minority over-sampling technique; Chawla et al., 2002) to balance outcomes; and (2) only those who accessed their digital intervention. For symptom-level change, models were repeated excluding baseline levels of modeled outcome (objective binge eating). Additional sensitivity analyses were conducted using intervention usage pattern data available from one RCT to see whether predictive accuracy improves.

3 | RESULTS

3.1 | Sample characteristics

Sample characteristics are presented in Table S3. Most participants were well-educated, White women who were highly symptomatic.

3.2 | Outcomes

Rate of uptake, adherence, and drop-out was 83%, 36%, and 57%, respectively. Objective binge eating decreased by an average of six episodes ($SD = 14$) from pre- to postintervention.

3.3 | Predicting engagement

The mean predictive performance for engagement outcomes across 100 iterations of validation data sets is presented in Table 1. Predictive performance was poor for the three engagement outcomes across the nine models based on AUC values.

3.3.1 | Predictor importance

Predictor importance was extracted from the SVM (linear) ML model. For uptake, these were motivation levels, age, overvaluation, eating concerns, and a current anxiety disorder. For adherence, these were age, depressive symptoms, current anxiety disorder, past anxiety disorder, and motivation levels. For dropout, these were age, education status, prior binge-eating disorder diagnosis, objective binge eating, and driven exercise (Table S4).

3.3.2 | Sensitivity analyses

When we balanced engagement outcomes by deriving an over-sampled training data using SMOTE technique (Table S5), predictive performance remained poor. Models on adherence and dropout that excluded "nonuptake" participants (Table S6) also had poor predictive performance.

3.4 | Predicting symptom change

Predictive performance based on amount of variance explained in symptom-level change (Table 2) was acceptable across the eight models. The elastic-net ($R^2 = .37$), SVM (linear) ($R^2 = .40$), SVM polynomial ($R^2 = .37$), and random forest models ($R^2 = .36$) explained slightly more variance in symptom change than the traditional linear model ($R^2 = .35$).

3.4.1 | Predictor importance

The five most important variables predictive of symptom change were baseline objective binge eating, subjective binge eating, eating concerns, self-induced vomiting, and overvaluation (Table S7). When removing baseline objective binge eating, predictive performance decreased. The CART model explained the most amount of variance in symptom change ($R^2 = .09$), slightly outperforming the traditional general linear model ($R^2 = .06$).

3.5 | Incorporating usage variables

Analyses were re-computed on two of participant groups for whom usage data were available. Analyses on this subsample were run with and without the addition of usage variables to enable comparison (Table S8). Incorporating usage variables produced strong predictive performance for adherence (AUC = 0.92–0.99) and dropout (AUC = 0.62–0.93) across all models except SVM (Radial). Predictive performance did not improve for symptom-level change.

TABLE 1   Mean model predictive performance of engagement from baseline variables across 100 iterations of training and testing data splits

				Negatives		Positives	
Modeling approach	Accuracy	AUC	F1 score	True negative	False positive	False negatives	True positives
Intervention uptake							
Generalized linear model	0.81	0.50	0.90	1.5%	98.5%	1.7%	98.3%
Elastic-net	0.82	0.50	0.90	0.2%	99.8%	0.3%	99.7%
SVM (Linear)	0.83	0.50	0.90	0.0%	100.0%	0.0%	100.0%
SVM (Polynomial)	0.82	0.50	0.90	0.0%	100.0%	0.0%	100.0%
SVM (Radial)	0.83	0.50	0.90	0.0%	100.0%	0.0%	100.0%
k-Nearest Neighbors	0.82	0.50	0.90	0.9%	99.1%	1.2%	98.8%
Random Forest	0.82	0.50	0.90	1.3%	98.7%	0.9%	99.1%
Classification/Regression Tree	0.79	0.50	0.88	4.6%	95.4%	4.9%	95.1%
Intervention adherence							
Generalized linear model	0.59	0.49	0.17	86.3%	13.7%	88.3%	11.7%
Elastic-net	0.63	0.50	0.09	97.7%	2.3%	98.3%	1.7%
SVM (Linear)	0.64	0.50	0.02	99.9%	0.1%	100%	0.0%
SVM (Polynomial)	0.64	0.50	0.03	99.8%	0.3%	99.8%	0.2%
SVM (Radial)	0.64	0.50	0.03	99.8%	0.2%	99.8%	0.2%
k-Nearest Neighbors	0.63	0.50	0.13	96.1%	3.9%	96.1%	3.9%
Random Forest	0.63	0.50	0.10	94.5%	5.5%	93.8%	6.2%
Classification/Regression Tree	0.59	0.50	0.30	80.1%	19.9%	79.8%	20.2%
Study dropout							
Generalized linear model	0.55	0.52	0.64	33.8%	66.2%	29.1%	70.9%
Elastic-net	0.57	0.51	0.71	7.1%	92.9%	4.9%	95.1%
SVM (Linear)	0.57	0.51	0.72	4.9%	95.1%	3.1%	96.9%
SVM (Polynomial)	0.57	0.51	0.71	7.2%	92.8%	5.6%	94.4%
SVM (Radial)	0.56	0.50	0.71	4.7%	95.3%	4.5%	95.5%
k-Nearest Neighbors	0.56	0.50	0.71	4.7%	95.3%	4.8%	95.2%
Random Forest	0.55	0.52	0.65	29.5%	70.5%	25.6%	74.4%
Classification/Regression Tree	0.55	0.52	0.66	27.2%	72.8%	22.9%	77.1%

Abbreviations: AUC, area under the curve; SVM, support vector machine.

TABLE 2   Mean model predictive performance for objective binge eating change from baseline variables across 100 iterations of training and testing data splits

	Binge eating change (including baseline binge eating)			Binge eating change (excluding baseline binge eating)		
Modeling approach	R^2	RMSE	MAE	R^2	RMSE	MAE
Generalized linear model	.35	11.64	8.31	.06	14.94	10.40
Elastic-net	.37	11.37	8.04	.06	14.64	10.10
SVM (Linear)	.40	11.21	7.63	.06	14.06	9.38
SVM (Polynomial)	.37	11.51	7.80	.05	14.16	9.22
SVM (Radial)	.29	12.24	8.22	.06	14.09	9.28
k-Nearest Neighbors	.15	13.35	8.92	.05	14.14	9.15
Random Forest	.36	11.51	7.88	.06	13.96	9.13
Classification and Regression Tree	.33	11.91	7.86	.09	13.97	9.21

Abbreviations: MAE, mean absolute error; RMSE, root-mean-square error; SVM, support vector machine.

4 | DISCUSSION

ML-based predictive performance from 36 baseline variables for engagement outcomes was poor and not superior to traditional logistic regression. ML models also explained less than 5% more variance than traditional generalized linear models on symptom-level change. Although some influential yet unanticipated predictors emerged, these are difficult to interpret given the poor model performance. Thus, a limited set of baseline predictors was not sufficient to detect a performance benefit of ML over traditional approaches, or that possible linear associations between modeled variables prevented interactive terms from enhancing predictive performance. Preliminary findings suggest that ML appears to not confer benefit to prediction modeling when the predictors chosen for analyses are small, unrelated or weakly related to outcomes. This finding is consistent with studies in depression (Lee et al., 2018), schizophrenia (Hettige et al., 2017), and anxiety (Wallert et al., 2018) reporting small to no incremental benefits of ML when restricting analyses to routinely collected baseline data, suggesting that the utility of ML may only emerge when multiple sources of data (sensor, neuroimaging, etc.) are modeled.

Study limitations must be considered. First, model validation was based on a subset of participants rather than an entirely new sample, which may contribute to overfitting concerns. While this has the advantage of holding methodological effects constant, testing on a new sample would enable stronger inferences. Moreover, model overfitting may be exasperated when small sample sizes are used in conjunction with a large number of predictors. Despite this, predictive performance was comparable across all ML models, including the elastic-net approach, which is useful even when the number of predictors is greater than the number of participants (Friedman et al., 2010). Second, model performance is dependent on the predictors analyzed. Our choice of baseline variables was guided mostly by pragmatic choices (to characterize the sample, those used as key outcome measures, etc.). Fluctuations in these or other variables closer to the point at which a person disengages may yield more precise prediction models, potentially explaining why prediction improved when incorporating usage variables in supplementary analyses. Third, our sample was mostly educated, White women and that the inclusion of predictors related to race, gender, and age may contribute to ongoing concerns about algorithmic bias within ML (Hooker, 2021). Future work utilizing ML in diverse samples is necessary.

This exploratory work offers avenues for future research. As it appears that our chosen predictors and timing of assessment are not decisive outcome determinants, future research should consider modeling the predictive performance of other sources of data that can be easily collected during digital intervention trials. Information on design choices, user experience, perceived usability, and reasons for disengagement may prove useful for generating accurate predictive models. Furthermore, assessing predictors repeatedly throughout the intervention phase may help to determine whether poor prediction is a function of measuring these constructs in a static manner and at an inappropriate time. Heterogeneity of treatment efficacy is well known (Linardon, Shatte, Messer, et al., 2020), such that two individuals with similar baseline scores may have markedly different symptom trajectories across different time intervals. Thus, it should not be surprising to find that earlier measurements of these constructs offer poor predictive value. Overall, findings will ideally encourage researchers to consider assessing these factors in future trials so that treatment outcome predictions can be optimized.

4.1 | Public significance statement

Being able to accurately predict patient outcomes is needed to accelerate the delivery of personalized eating disorder treatments. We sought to generate accurate predictive models from routinely collected baseline data from digital interventions using data-driven machine learning techniques. Our findings highlight the pitfalls of relying solely on baseline information for predicting success from digital interventions, instead suggesting that other sources of big data may prove more useful.

AUTHOR CONTRIBUTIONS

Jake Linardon: Conceptualization; funding acquisition; investigation; methodology; writing – original draft; writing – review and editing. **Matthew Fuller-Tyszkiewicz:** Conceptualization; data curation; supervision; writing – original draft; writing – review and editing. **Adrian Shatte:** Conceptualization; data curation; methodology; software; supervision; writing – review and editing. **Christopher Greenwood:** Conceptualization; data curation; formal analysis; investigation; methodology; writing – original draft; writing – review and editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Please contact the corresponding author to gain access to de-identified data if necessary.

ORCID

Jake Linardon https://orcid.org/0000-0003-4475-7139

REFERENCES

Bremer, V., Chow, P. I., Funk, B., Thorndike, F. P., & Ritterband, L. M. (2020). Developing a process for the analysis of user journeys and the prediction of dropout in digital health interventions: Machine learning approach. *Journal of Medical Internet Research*, *22*(10), e17738. https://doi.org/10.2196/17738

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R.,

Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, 20(2), 154–170. https://doi.org/10.1002/wps.20882

Chien, I., Enrique, A., Palacios, J., Regan, T., Keegan, D., Carter, D., Tschiatschek, S., Nori, A., Thieme, A., Richards, D., Doherty, G., & Belgrave, D. (2020). A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. *JAMA Network Open*, 3(7), e2010791. https://doi.org/10.1001/jamanetworkopen.2020.10791

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, 110, 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

Fitzsimmons-Craft, E. E., Taylor, C. B., Graham, A. K., Sadeh-Sharvit, S., Balantekin, K. N., Eichen, D. M., Monterubio, G. E., Goel, N. J., Flatt, R. E., Karam, A. M., Firebaugh, M. L., Jacobi, C., Jo, B., Trockel, M. T., & Wilfley, D. E. (2020). Effectiveness of a digital cognitive behavior therapy–guided self-help intervention for eating disorders in college women: A cluster randomized clinical trial. *JAMA Network Open*, 3(8), e2015633. https://doi.org/10.1001/jamanetworkopen.2020.15633

Flygare, O., Enander, J., Andersson, E., Ljótsson, B., Ivanov, V. Z., Mataix-Cols, D., & Rück, C. (2020). Predictors of remission from body dysmorphic disorder after internet-delivered cognitive behavior therapy: A machine learning approach. *BMC Psychiatry*, 20(1), 1–9. https://doi.org/10.1186/s12888-020-02655-4

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.

Hettige, N. C., Nguyen, T. B., Yuan, C., Rajakulendran, T., Baddour, J., Bhagwat, N., Bani-Fatemi, A., Voineskos, A. N., Mallar Chakravarty, M., & De Luca, V. (2017). Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach. *General Hospital Psychiatry*, 47, 20–28. https://doi.org/10.1016/j.genhosppsych.2017.03.001

Hooker, S. (2021). Moving beyond "algorithmic bias is a data problem". *Patterns*, 2(4), 100241.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kuhn. (2021). caret: Classification and regression training. R package version 6.0-88. https://CRAN.R-project.org/package=caret

Lee, Y., Ragguett, R.-M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A., Brietzke, E., Lin, K., Pan, Z., Subramaniapillai, M., Chan, T. C. Y., Fus, D., Park, C., Musial, N., Zuckerman, H., Chen, V. C., Ho, R., Rong, C., & McIntyre, R. (2018). Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review. *Journal of Affective Disorders*, 241, 519–532. https://doi.org/10.1016/j.jad.2018.08.073

Lekkas, D., Price, G., McFadden, J., & Jacobson, N. C. (2021). The application of machine learning to online mindfulness intervention data: A primer and empirical example in compliance assessment. *Mindfulness*, 12(10), 2519–2534. https://doi.org/10.1007/s12671-021-01723-4

Linardon, J., Messer, M., Shatte, A., Greenwood, C. J., Rosato, J., Rathgen, A., Skvarc, D., & Fuller-Tyszkiewicz, M. (2021). Does the method of content delivery matter? Randomized controlled comparison of an internet-based intervention for eating disorder symptoms with and without interactive functionality. *Behavior Therapy*, 53, 508–520. https://doi.org/10.1016/j.beth.2021.12.001

Linardon, J., Messer, M., Shatte, A., Skvarc, D., Rosato, J., Rathgen, A., & Fuller-Tyszkiewicz, M. (2021). Targeting dietary restraint to reduce binge eating: A randomised controlled trial of a blended internet- and smartphone app-based intervention. *Psychological Medicine*, 1-11, 1–11. https://doi.org/10.1017/S0033291721002786

Linardon, J., Rosato, J., & Messer, M. (2020). Break binge eating: Reach, engagement, and user profile of an internet-based psychoeducational and self-help platform for eating disorders. *International Journal of Eating Disorders*, 53, 1719–1728. https://doi.org/10.1002/eat.23356

Linardon, J., Shatte, A., Messer, M., Firth, J., & Fuller-Tyszkiewicz, M. (2020). E-mental health interventions for the treatment and prevention of eating disorders: An updated systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*, 88, 994–1007. https://doi.org/10.1037/ccp0000575

Linardon, J., Shatte, A., Rosato, J., & Fuller-Tyszkiewicz, M. (2020). Efficacy of a transdiagnostic cognitive-behavioral intervention for eating disorder psychopathology delivered through a smartphone app: A randomized controlled trial. *Psychological Medicine*, 1-12, 1–12. https://doi.org/10.1017/S0033291720003426

Pearson, R., Pisner, D., Meyer, B., Shumake, J., & Beevers, C. G. (2019). A machine learning ensemble to predict treatment outcomes following an internet intervention for depression. *Psychological Medicine*, 49(14), 2330–2341. https://doi.org/10.1017/S003329171800315X

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.

Wallert, J., Gustafson, E., Held, C., Madison, G., Norlund, F., von Essen, L., & Olsson, E. M. G. (2018). Predicting adherence to internet-delivered psychotherapy for symptoms of depression and anxiety after myocardial infarction: Machine learning insights from the U-CARE heart randomized controlled trial. *Journal of Medical Internet Research*, 20(10), e10754. https://doi.org/10.2196/10754

Weissman, R. S., & Rosselli, F. (2017). Reducing the burden of suffering from eating disorders: Unmet treatment needs, cost of illness, and the quest for cost-effectiveness. *Behaviour Research and Therapy*, 88, 49–64. https://doi.org/10.1016/j.brat.2016.09.006

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.