Revised and resubmitted to Clinical Psychological Science

The development and internal evaluation of a predictive model to identify for whom Mindfulness-Based Cognitive Therapy offers superior relapse prevention for recurrent depression versus maintenance antidepressant medication

Zachary D. Cohen, PhD[1]

Robert J. DeRubeis, Professor[2]

Rachel Hayes, PhD[3]

Edward. R. Watkins, Professor[4]

Glyn Lewis, Professor[5]

Richard Byng, Professor[5,6]

Sarah Byford, Professor[7]

Catherine Crane, PhD[8]

Willem Kuyken, Professor[8]

Tim Dalgleish, Professor[9,10]*†

Susanne Schweizer, Professor[9, 11]*

[1]Department of Psychiatry, University of California Los Angeles, Los Angeles, USA

[2]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]University of Exeter Medical School, South Cloisters, St Luke's Campus, Exeter, UK

[4]Sir Henry Wellcome Mood Disorder Center, University of Exeter, Exeter, UK

[5]Community Primary Care Research Group, University of Plymouth, Plymouth, UK

[6]National Institute of Health Research Collaboration for Leadership in Applied Health Research and Care, South West Peninsula

[7]Health Service and Population Research Department, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

[8]Department of Psychiatry, Medical Sciences Division, University of Oxford, Oxford, UK

[9]Department of Psychology, University of Cambridge, Cambridge, UK

[10]Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK

[11]School of Psychology, University of New South Wales, Sydney, Australia

[?]Division of Psychiatry, Faulty of Brain Sciences, University College London, London, UK

*These authors should be regarded as Joint Senior Authors.

† Corresponding authors: Zachary Cohen, zachary.d.cohen@gmail.com; Tim Dalgleish, MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, CB2 7EF, Cambridge, UK; Susanne Schweizer, Department of Psychology, University of Cambridge, Downing Site, CB2 3EB, Cambridge, UK.

Acknowledgements:

This research was funded by the National Institute for Health Research Health Technology Assessment (NIHR HTA) programme (project number: 08/56/01). The views expressed in this publication are those of the authors and do not necessarily reflect those of the HTA programme, NIHR, NHS, the Department of Health and Social Care, the Wellcome Trust or MQ.

Tim Dalgleish was partly supported by the National Institute for Health Research Cambridge Biomedical Research Centre and the UK Medical Research Council

(SUAG/043 G101400). Susanne Schweizer was supported by the Wellcome Trust (209127/Z/17/Z). Zachary D. Cohen and Robert J. DeRubeis were partly supported by MQ (MQ14PM_27). Willem Kuyken and Catherine Crane were supported by the Wellcome Trust [104908/Z/ 14/Z and 107496/Z/15/Z]. Richard Byng was supported by the NIHR Collaboration for Leadership in Applied Health Research and Care South West Peninsula at the Royal Devon and Exeter NHS Foundation Trust. The opinions and assertions contained in this article should not be construed as reflecting the views of the sponsors. We are grateful to the participants for their time in taking part in this trial and we also thank the following colleagues who have contributed to the PREVENT study, through recruitment, retention and treatment of patients or provision of administrative support: Claire Brejcha, Jessica Cardy, Aaron Causley, Suzanne Cowderoy, Alison Evans, Felix Gradinger, Surinder Kaur, Jonathan Richards, Harry Sutton, Rachael Vicary, Alice Weaver, Jenny Wilks and Matthew Williams.

Abstract

Major depressive disorder is highly recurrent over the lifespan, even following successful pharmacological and/or psychological intervention. We aimed to develop clinical prediction models that could be used to optimize treatment recommendations, when choosing between continuing antidepressant medication (ADM) or switching to Mindfulness-Based Cognitive Therapy (MBCT) for individuals with recurrent depression. Using data from the PREVENT trial (*N*=424; Kuyken et al., 2015), we constructed multivariable prognostic models combining demographic, clinical and psychological factors using elastic net regularized regression to predict relapse over 24-month follow-up in the ADM and MBCT groups. Only the ADM-model's discrimination performance (AUC=.68) was superior to a benchmark comparison model based on baseline depression severity (one-tailed DeLong's test; Z=2.8, *P*=.003). Among the third of the sample with the poorest ADM prognoses, relapse rates (MBCT=48%, ADM=70%) and survival times (*z*=−2.7; *P*=.008) were superior for those who switched to MBCT versus those who continued ADM. For those with moderate-to-good ADM prognosis, both treatments resulted in similar likelihood of relapse. If replicated, the results suggest that predictive modeling could inform clinical decision-making around relapse prevention following full or partial remission to antidepressants in recurrent depression.

Globally, depressive disorders are now the leading causes of life years lived with disability (Patel et al., 2016, 2018). In many cases the course of depression is recurrent over the lifespan (Kessler & Bromet, 2013), even following successful acute-phase interventions (Cuijpers et al., 2013). Successful prevention of the return of depression is therefore key to alleviating the individual and societal burden of depressive disorders. Antidepressant medication (ADM) following successful treatment is currently the predominant preventive intervention targeted at depressive relapse[1]. Multiple agencies, including the UK National Institute for Health and Care Excellence (NICE), the British Association for Pharmacology (Cleare et al., 2015) and the American Psychiatric Association, recommend prescription of ADM after remission if a person is deemed at high risk of relapse because of multiple previous episodes or high residual symptoms (Gelenberg et al., 2010; NICE, 2009). An international review of 13 sets of ADM guidelines revealed that recommendations for the duration of such continuation or maintenance[2] treatment in those deemed at high risk ranged from 1 year to lifelong or indefinite (Piek et al., 2010). Unsurprisingly, therefore, longer-term use of ADMs is high and rising (Mojtabai & Olfson, 2014; OECD, 2013), accounting for the recorded increase in person-years on ADMs from 0.73 in 1995 to 4.94 in 2012 reported in the UK (McCrea et al., 2016).

[1] We note that a recent systematic review of 56 studies and almost 40K subjects by De Zwart and colleagues (2019) re-examined the distinction between relapse and recurrence established by Frank et al. (1991) and concluded that "the idea that a reoccurrence of depressive symptoms shortly after their initial remission constitutes a 'relapse' of the previous episode, whereas their later reoccurrence is the first sign of an entirely new episode, is a model that lacks empirical support" (De Zwart et al., 2019, p. 544). Therefore, in this paper, we will use the term *relapse* to describe a return of depressive symptoms meeting DSM-IV criteria (as assessed via Structured Clinical Interview for DSM-IV), regardless of when during 24-month follow-up it occurred.

[2] We will use continuation and maintenance interchangeably, but note that in some contexts (e.g., DeRubeis et al., 2019), these terms are used more specifically to differentiate treatment following remission and up to recovery (continuation) from treatment past the point of recovery (maintenance).

The anti-depressant benefits of longer-term ADM use are tempered by diverse physical and emotional side effects in the majority of patients (Bet et al., 2013; Cartwright et al., 2016), tachyphylaxis and other loss of response phenomena (Bosman et al., 2018; Fornaro et al., 2019; Kinrys et al., 2019), and user surveys indicating a desire for evidence-based psychosocial interventions as an alternative to ADMs for all aspects of depression management (Dorow et al., 2018; McHugh et al., 2013; Schweizer et al., 2010). Treatment guidelines stipulate that such patient preferences should inform treatment selection through a process of shared decision making (Weston, 2001) and there is some evidence that treatment outcomes are superior for preferred versus non-preferred treatments (Windle et al., 2019; Kwan et al., 2010; Shay & Lafata, 2015).

A critical component of effective shared decision making is ensuring that comparative evidence for different interventions in the context of the patient's own clinical profile – *what works for whom* – is available at the point of care delivery (Winston et al., 2018). This information can come in a variety of different forms, including decision aids (Stacey et al., 2017), or more quantitative outcomes from clinical prediction models (Bonnett et al., 2019). Recent methodological and empirical advances in 'precision medicine' (Collins & Varmus, 2015) have generated prediction models that provide indices to identify which patients might expect improved clinical outcomes following different acute treatments for depression (Cohen et al., 2021; Chekroud et al., 2021; Cohen & DeRubeis, 2018). Here, we focus on identifying patient characteristics that could putatively guide the treatment choice between continuing ADM versus a psychosocial alternative intervention – Mindfulness-Based Cognitive Therapy (MBCT) with support to taper or discontinue antidepressant treatment – for the prevention of depressive relapse.

MBCT is an 8-week group-based programme that has emerged as a leading evidence-based psychological intervention for relapse prevention in recurrent depression (Kuyken et al., 2016). In a recent multicentre definitive randomised controlled trial (*N*= 424) – the PREVENT trial (Kuyken, Byford, et al., 2010; Kuyken et al., 2014) – we evaluated MBCT combined with support to taper or discontinue ADM against maintenance of a clinical dose of ADM for two years in patients (age≥18) with recurrent depression (at least 3 previous episodes) who were in partial or full remission on ADM. The trial showed no significant differences in relapse over 2-years between the MBCT and ADM groups (hazard ratio: 0.89, 95% CI 0.67–1.18; p=0.43; relapse rate MBCT 44% vs. 47% ADM; Kuyken et al., 2015), a finding corroborated by an individual patient data meta-analysis of 1258 patients from nine RCTs (Kuyken et al., 2016). Recent meta-analytic work has confirmed and expanded these findings: a network meta-analysis by McCartney et al. (2021) provided additional evidence that MBCT is superior to control conditions in terms of rate of relapse (MBCT vs treatment as usual) or time to relapse (MBCT vs treatment as usual or placebo), and a meta-analysis by Breedvelt and colleagues (2020) added evidence of the superiority of combination psychological prevention and continuation ADM over continuation ADM alone. The participants in PREVENT were assessed at trial baseline on a broad range of psychosocial variables that putatively have a bearing on treatment outcome (Kuyken et al., 2015). Here we explored whether we could generate a multivariable predictive algorithm using baseline patient characteristics and outcome data from PREVENT that reliably differentiates patients who will benefit from switching to MBCT and tapering their ADM from those for whom continuation of ADM likely provides the best option for the prevention of depression relapse.

METHODS

Dataset description

The full PREVENT sample comprised 424 individuals randomized (1:1) to ADM or MBCT. Of these, participants with >20% missing data on predictor variables (*n*=15), no data beyond baseline (*n*=17), or not in receipt of a dose of MBCT deemed sufficient (at least 4 sessions; following the PREVENT trial protocol: Kuyken et al., 2010, 2015) for evaluation of MBCT as an intervention alternative (*n*=25), were excluded from the primary analyses. This led to a sample of 367 participants for the primary analyses. The data exclusion pipeline is shown in Figure S1 in the Supplementary Materials. Sensitivity analyses were performed to probe the impact of removing the (*n*=25) who were excluded due to inadequate MBCT dose. The results for this larger sample (*n*=392) are included in the Supplementary Materials. Descriptive data for the predictor variables at baseline are provided in the Supplementary Materials, as are comparisons of the various groups (ADM vs. MBCT, excluded vs analysis sample). These comparisons indicated that there was a significantly greater proportion of women in the ADM group, and ADM participants reported more comorbid diagnoses and had a lower probability that their most recent episode of depression was chronic (>24 months in duration), at baseline, compared to the MBCT group (Table S#). Relative to the analysis sample, excluded participants reported significantly more comorbidities, were on average five years younger, and reported lower feelings of isolation on the Self Compassion Scale (Table S#).

Predictor variables

The PREVENT study included a wide range of 53 potential demographic, clinical and psychological predictor variables (Supplemental Table #). The demographic and clinical predictors were selected because they are available in clinical practice, and indeed many are commonly included as part of routine diagnostic procedures. Psychological

predictors included standardized self-report measures of potential mechanisms of treatment efficacy (including mindfulness, self- and other-compassion and repetitive thinking). Details are presented in Table S2 in the Supplementary Materials.

Missing predictor variable data at baseline were imputed using the full (*n*=424) sample via the missForest (Stekhoven & Bühlmann, 2012) package in R (R Core Team, 2013), which implements a random forest-based non-parametric imputation approach. Random forest-based imputation compared favorably in several evaluations of different imputation approaches (Stekhoven & Bühlmann, 2012; Waljee et al., 2013; Shah et al., 2014; c.f., Hong & Lynn, 2020). Additional information regarding imputation can be found in the supplemental materials.

For the 53 potential predictors assessed at baseline in the PREVENT data, following imputation, continuous variables were z-scored (standardized to have SD=1 and mean=0) and binary variables were set to -0.5 and 0.5.

Statistical approach to treatment selection

An in-depth description of creating and evaluating treatment recommendations can be found in a recent review of treatment selection (Cohen & DeRubeis, 2018). The core concept is that statistical models are constructed and used to generate predictions for an individual in two (or more) treatments, and then those predictions are used to determine which treatment to pursue (Cohen, Ashar, & DeRubeis, 2019). Much of the work in this space (e.g., the Personalized Advantage Index [PAI] approach; DeRubeis et al., 2014) has been based on the proportional interaction model. Luedtke and colleagues (2019) highlighted potential problems with the use of this approach, whose implicit estimation and testing of interaction effects (versus main effects) requires larger samples, in the small RCT samples that are often available. Their simulation work suggested that sample sizes of at least 300 per condition are required for adequate statistical power to detect clinically

significant improvements in response associated with model-based treatment selection. Other approaches that have been demonstrated rely solely on prognostic models (e.g., Lorenzo-Luaces et al., 2017; Wiltsey-Stirman et al., 2021). For a discussion and contrasting of these different approaches, see Cohen et al. (2021). Following the approach proposed by Kessler et al. (2017) and demonstrated by Deisenhofer and colleagues (2018), we constructed separate prognostic algorithms for each of our two treatment conditions (MBCT and ADM). For each patient, a 'factual prediction' – how well that patient was expected do in their actual treatment arm based on their scores on the variables selected for that treatment's prognostic model – was generated, along with a 'counterfactual prediction' – how well the patient would hypothetically have done in the alternative treatment arm based on their scores on the predictors that were included in the prognostic model for the alternative treatment arm.

In this approach, the predictive performance of each of the two separate treatment arm algorithms could be independently evaluated (see below for information about cross-validation) using the factual predictions and the observed outcomes. In the event that both algorithms had yielded inaccurate factual predictions, this would have revealed that the data did not provide a useful signal for prediction purposes. In the event that both models yielded accurate factual predictions, the computed difference between the sets of predictions for MBCT and ADM could have served as an index for each patient, indicating which of the two treatments would be optimal (Cohen & DeRubeis, 2018). Finally, if only one of the models (e.g., "TxA") yielded accurate factual predictions, that model on its own could be evaluated for its potential utility for guiding treatment decisions. Patients could be arrayed based on their predicted outcome in the condition with the reliable prognostic model (TxA). In the absence reliable information about expected response to the other treatment ("TxB"), those with poor prognoses in TxA

could be reasonably advised to try TxB, whereas a sensible recommendation for those with good prognoses in TxA would be, unsurprisingly, TxA. Thus, the expectation in this scenario would be that differential response would be observed across the spectrum of TxA prognosis: Among those with poorer prognoses in TxA, outcomes, on average, for those who received TxA would be worse than for those who received TxB, and vice versa for those with better prognoses in TxA.

We applied this approach to the PREVENT data, and below we outline the steps of variable selection, cross-validation, and assessment of model fit, involved in building and evaluating the prognostic algorithms for MBCT and ADM. Although analyses revealed the MBCT model to have poor predictive performance (as indicated by low AUC), the ADM model evidenced good predictive performance and was superior to a benchmark model constructed only using baseline depression severity. Consequently, we generated and evaluated the treatment selection indices based on the ADM prognostic model only. This allowed us to ask the question of whether there were differential outcomes for those who received MBCT and ADM in patients predicted to do well, moderately, or poorly if they continued with ADM, which would provide evidence for the potential utility of the ADM model for guiding treatment selection. The details regarding the model building and evaluation for the poorly fitting MBCT model are described in detail in the Supplementary Materials.

Cross-validation.

When using cross-validation in the context of predictive model evaluation it is essential to protect against ‘double-dipping’ (Hastie et al., 2009). For example, it is critical that the predictions that are evaluated are generated from models that are

constructed (in terms of variable selection, hyperparameter tuning, and weight setting) without the use of the individuals for whom the predictions are being made.

We performed 10-fold cross-validation (Hastie et al., 2009), which involved splitting both the ADM and MBCT samples into 10 sub-groups, balanced on outcomes. Each of the 10 ADM sub-groups were then held-out, and a prediction model was constructed using the remaining 9 ADM sub-groups as the training sample. That model was then applied to the 10th ADM group to generate factual predictions of expected response in ADM, and was also applied to the entire MBCT sample to generate *counterfactual* predictions of their expected response if they had received ADM. The protections needed differ when generating factual and counterfactual prediction for each treatment arm. When predicting ADM outcomes for the MBCT sample, no cross-validation is needed, as the ADM model was constructed without the MBCT individuals and thus can be applied to these individuals without concern over so-called "double-dipping." This process was repeated 9 more times for each of the other 9 ADM sub-groups, resulting in the generation of a single "protected" factual prediction for each of the individuals in the ADM condition, and 10 protected counterfactual predictions (one from each of the 10 ADM models) for each of the individuals in the MBCT condition, which were then averaged to create an ensemble counterfactual prediction of how those who received MBCT would have been expected to fare had they received ADM. The analogous process was then performed for the MBCT group, resulting in each individual in MBCT receiving a single factual prediction of their outcomes in MBCT and those in the ADM condition receiving ensemble counterfactuals for their expected outcomes had they received MBCT. Finally, in order to provide a benchmark to help in the evaluation of these multivariable prediction models, we used the same cross-validation strategy, again in both groups, to generate predictions from "severity only" models (constructed using

logistic regression), in which the only predictor available to the models was baseline symptom severity on the clinician-assessed Hamilton Rating Scale for Depression (HAMD; Hamilton, 1967) (Figure S1), assessed using the 17-item GRID-HAMD (Williams et al., 2018). See Figure # for a schematic describing the entire analytic pipeline.

Modeling Via Elastic Net Regularized Regression.

Multivariable prognostic models were constructed using elastic net regularized regression (ENRR) (Zou & Hastie, 2005). ENRR allows for the estimation of the predictive utility of a large number of variables and its use has been demonstrated and extensively discussed in several previous predictive modeling efforts (Webb et al., 2020; Buckman et al., 2021; Chekroud et al., 2016, 2017; Iniesta et al., 2016; Kim et al., 2019; Cohen et al., 2019). ENRR combines the L1 and L2 penalization, providing a hybrid of LASSO and Ridge regression, thus addressing issues of correlated predictors and over-fitting by shrinking coefficients of correlated predictors towards each other, and by removing uninformative predictors from the model (Hastie et al., 2009). ENRR was implemented using the R package glmnet (Friedman et al., 2010). To reduce a source of potential bias (risk of overfitting due to information leakage from the test cases; Pearson et al., 2018) that can arise when a grid search is performed for hyperparameter setting in the context of cross-validation, we used three tuning loops (Zou & Hastie, 2005), 10-fold cross-validation (Friedman et al., 2010; Zou & Hastie 2005, p.310), and a small set of alpha values (.01, .5, .99) recommended by Shumake and colleagues' BeSet Package (cite). Additionally, the regularization parameter lambda was selected using the one-standard-error-rule, which helps to avoid overfitting and elevated Type I error (James et al., 2013; Waldmann, et al., 2013).

Table 1. Descriptive data for the predictor variables at baseline (primary analysis sample N=367)

Predictor	ADM (*n* = 195)	MBCT (*n* = 172)	Continuous: Mean difference Categorical: X^2	*P* Value
Demographic characteristics				
Age (years)			2.52	.05
Mean (*sd*)	48.77 (12.69)	51.30 (11.56)		
Range	20-79	24-78		
Female (%)	160 (82)	118 (69)	8.28	.004
Education			11.30	.33
No Qualification	10 (5)	10 (6)		
GCSE	38 (19)	24 (14)		
UK Advanced Level	27 (14)	17 (10)		
Vocational qualification	65 (33)	56 (32)		
Bachelor's degree	32 (17)	44 (26)		
Master's degree	9 (5)	9 (5)		

Professional training/PhD	14 (7)	12 (7)		
Relationship			0.33	.57
Yes (Married/Civil partnership/Cohabiting)	128 (66)	107 (62)		
No (Single/Divorced/Widowed)	67 (34)	65 (38)		
Employed (full- or part-time vs. unemployed)	119 (61)	98 (57)	0.46	.50
Clinical characteristics				
Clinician-rated depressive symptoms (HAMD)			0.14	.75
Mean (*sd*)	4.62 (4.31)	4.76 (4.27)		
Range	0-20	0-19		
Self-reported depressive symptoms (BDI-II)			–0.83	.43
Mean (*sd*)	14.39 (10.08)	13.55 (10.14)		
Range	0-42	0-42		
Age of onset			–0.21	.86
Mean (*sd*)	25.06 (11.96)	25.84 (11.60)		
Range	6-56	5-56		

Chronicity (previous episode ≥24 months)	38 (19)	53 (31)	5.69	.02
Previous psychological treatment	100 (51)	84 (49)	0.13	.72
Previous suicide attempt	49 (25)	33 (19)	1.53	.22
Family history of depression	98 (50)	90 (52)	0.08	.77
Number of comorbid diagnoses			–0.21	.006
Mean (*sd*)	0.61 (0.77)	0.40 (0.65)		
Range	0-2	0-2		
Psychological mechanisms				
Five-Facets Mindfulness Questionnaire (FFMQ)				
Observe			0.14	.81
M (*sd*)	24.10 (5.62)	24.25 (5.67)		
Range	11-37	8-39		
Describe			0.31	.66
M (*sd*)	26.04 (7.15)	26.35 (6.62)		
Range	8-40	10-40		

Aware			0.26	.65
M (*sd*)	23.94 (5.15)	24.20 (5.58)		
Range	10-37	10-37		
Non-Judging			0.16	.82
M (*sd*)	24.86 (6.34)	25.01 (6.93)		
Range	10-40	8-39		
Non-Reactivity			0.83	.11
M (*sd*)	19.28 (4.59)	20.08 (5.93)		
Range	10-31	7-32		
Self-Compassion Scale (SCS)				
Self-Kindness			0.08	.84
M (*sd*)	12.60 (3.90)	12.52 (4.17)		
Range	5-22	5-22		
Self-Judgement			0.12	.76
M (*sd*)	11.76 (3.77)	11.88 (3.91)		

Range	5-21	5-21		
Common Humanity			–0.23	.57
M (*sd*)	11.82 (3.79)	11.60 (3.85)		
Range	4-20	4-20		
Isolation			–0.23	.50
M (*sd*)	9.56 (3.17)	9.33 (3.30)		
Range	4-17	4-17		
Mindfulness			–0.03	.94
M (*sd*)	11.81 (3.12)	11.79 (3.38)		
Range	4-20	4-20		
Over-Identification			0.07	.84
M (*sd*)	9.28 (3.05)	9.31 (3.28)		
Range	4-17	4-17		
Compassion for others			0.21	.54
M (*sd*)	9.17 (3.30)	9.38 (3.35)		

Range	4-17	4-22		
Dispositional Positive Emotions Scale (DPES)				
Joy			–0.47	.30
M (*sd*)	17.21 (4.30)	16.74 (4.38)		
Range	6-28	6-28		
Contentment			–0.12	.78
M (*sd*)	14.17 (3.92)	14.05 (4.15)		
Range	5-25	5-25		
Love			–0.29	.53
M (*sd*)	19.22 (4.62)	18.92 (4.23)		
Range	8-30	9-29		
Compassion			–0.20	.53
M (*sd*)	21.36 (2.99)	21.16 (3.22)		
Range	12-25	13-25		
Awe			0.53	.23

M (*sd*)	19.25 (4.44)	19.78 (4.09)		
Range	8-30	9-30		
Curiosity			0.60	.09
M (*sd*)	20.45 (3.19)	21.05 (3.55)		
Range	7-30	12-30		
Cognitive Emotion Regulation Questionnaire (CERQ)				
Catastrophizing			0.45	.18
M (*sd*)	8.54 (2.94)	8.99 (3.39)		
Range	4-18	4-18		
Rumination			0.16	.65
M (*sd*)	12.04 (3.54)	12.20 (3.42)		
Range	4-20	5-20		
Other-blame			0.33	.22
M (*sd*)	7.37 (2.29)	7.70 (2.85)		

Range	4-15	4-15		
Self-blame			0.34	.36
M (*sd*)	10.72 (3.38)	11.05 (3.64)		
Range	4-20	4-20		
Acceptance			–0.02	.95
M (*sd*)	11.76 (3.01)	11.74 (3.03)		
Range	6-20	5-19		
Positive Refocusing			–0.51	.11
M (*sd*)	8.04 (3.01)	7.53 (3.09)		
Range	4-15	4-15		
Positive Reappraisal			0.19	.64
M (*sd*)	10.00 (3.68)	10.19 (4.07)		
Range	4-18	4-18		
Putting into Perspective			–0.11	.77
M (*sd*)	10.94 (3.40)	10.83 (3.89)		

Range	4-20	4-20		
Refocus on Planning			0.24	.52
M (*sd*)	10.63 (3.31)	10.87 (3.71)		
Range	4-20	4-20		
Cambridge-Exeter Repetitive Thought Scale (CERTS)				
Negative Rumination			0.71	.65
M (*sd*)	73.74 (14.71)	74.18 (14.78)		
Range	38-100	38-100		
Positive Rumination			0.77	.19
M (*sd*)	22.50 (5.60)	23.27 (5.63)		
Range	10-36	10-36		
Constructive Rumination			0.44	.18
M (*sd*)	10.92 (3.03)	11.35 (3.09)		
Range	4-20	4-20		

Unresolution			–0.00	.99
M (*sd*)	12.35 (2.55)	12.34 (2.77)		
Range	7-18	7-18		
Moving On			–0.17	.35
M (*sd*)	7.85 (1.69)	7.68 (1.74)		
Range	4-12	4-12		
Other				
Childhood abuse (MOPS)			–0.47	.49
High	101 (51)	82 (48)		
Low	94 (49)	90 (52)		
Self-Efficacy (GSS)			–0.02	.98
Mean (*sd*)	32.29 (7.74)	32.27 (8.18)		
Range	13-50	10-50		
Stigmatisation (SN)			–0.32	.65
Mean (*sd*)	20.88 (6.36)	20.56 (7.10)		

Range	7-35	7-35		
Recognizing warning signs (WS)			0.75	.21
Mean (*sd*)	18.26 (5.26)	19.01 (5.55)		
Range	6-30	6-30		
Relationship Satisfaction (RS)			0.62	.36
Mean (*sd*)	26.45 (6.66)	27.06 (6.19)		
Range	7-35	9-35		
Preference for MBCT			–0.02	.72
	4.51 (0.65)	4.49 (0.68)		
	3-5	3-5		
Preference for ADM			0.05	.63
	3.10 (1.09)	3.16 (1.09)		
	1-5	1-5		
Preference for Therapy Type			–0.10	.37
	1.80 (1.06)	1.70 (1.07)		

1-5 1-5

Table 1. HAMD = Hamilton Depression Rating Scale (Hamilton, 1967), assessed using the 17-item GRID-HAMD (Williams et al., 2018); BDI-II = Beck Depression Inventory Version-II (Beck et al., 1996); MOPS = Measure of Parenting Style (Parker et al., 1997); GSS = General Self-Efficacy Scale (Sherer et al., 1982); SN = Stigmatization and Normalization (bespoke questions); WS = Warning signs (bespoke questions); RS= Relationship satisfaction (bespoke questions). Individuals were asked to complete all measures (except for the MOPS) with respect to the previous two weeks. All scales except for the HAMD, BDI-II, and MOPS were scored on a 5-point Likert scale irrespective of their original scoring range. The scaling was standardized to facilitate interpretation from factor analyses and similar computations planned for the trial. The labels of the original scales were maintained. Further details on the psychological predictor variables are presented in Table S2 in the Supplementary Materials.

Evaluating the model. The model was evaluated using receiver operating characteristic (ROC) curves, which delineate the relative sensitivity (true positive rate) and specificity (false positive rate) of a model's predictions at different thresholds. The area under the curve (AUC) of the ROC, which typically ranges between 0.5 for chance models and 1 for models with perfect predictions, is then used to summarize a given model's predictive power. In this context, the AUC is equivalent to the calibration or c-statistic. We computed AUCs for the ENRR models' factual predictions for patients in each treatment arm (ADM and MBCT) and compared the AUCs using the DeLong test, one-tailed as our hypothesis was that the multivariable models would outperform the benchmark models. We also computed the AUC for each treatment arm for the depression-severity-at-baseline-only logistic regression models (HAMD) in each arm as a benchmark for the more complex multivariable models.

Evaluating prognostic utility. As noted**,** the prediction model for MBCT had an unacceptable level of agreement, no better than chance and not superior to the HAMD model (see Supplementary Materials) for prediction purposes and so we evaluated prognostic utility based on the ADM model only. The rationale therefore was to evaluate whether those who are predicted to have a high risk of relapse if they continue ADM might have a better predicted outcome with a switch to MBCT. Similarly, we wanted to examine whether those predicted to have a good prognosis with ADM might be better advised to continue the treatment regimen they are already following, namely ADM.

To evaluate the overall utility of the predictions generated by the ADM prognostic model in guiding treatment selection, we used two tertiles to divide the sample into three groups (Altman & Bland, 1994) based on risk of relapse in ADM (good ADM prognosis, moderate ADM prognosis, and poor ADM prognosis). Sample sizes and descriptive statistics for the ADM prognoses (i.e., means, standard deviations, and ranges) for three groups, broken down by

treatment received, can be found in the supplement. Predictive utility of the ADM prognostic index was then evaluated by examining the time-to-relapse (in a survival analysis using Cox regression), as well as overall relapse rates, over the two-year follow-up with treatment condition (ADM, MBCT), ADM-prognosis (both as a continuous variable and in categorical form: good, moderate, poor), and their interaction as factors. For any significant interactions the effects of treatment group were analysed within each prognostic category.

RESULTS

Model predicting relapse in the ADM treatment arm

Data for the poorly fitting MBCT model (AUC = 0.51) are presented in the Supplementary Materials and Figure S2.

Details on the MBCT prognostic models are provided in the Supplementary Materials. Following variable section and weighting, and cross-validation, the ADM multi-variable prediction model (henceforth simply called the ADM model) included the following eight baseline variables (from our set of 53) as predictors of relapse over the course of the 24-month follow-up period: Presence of previous suicide attempt(s) + Level of self-reported childhood abuse on the MOPS + Employment status + Previous depressive episode chronicity + Self-efficacy + Use of acceptance as an emotion regulation strategy + Dispositional joy + Ruminative unresolution. The predictor weightings are presented in Table 2. Care must be taken when seeking to interpret the relationship of any one variable to outcome when that variable is part of a multi-variable model. However, for an illustration of the degree and direction of association between the individual variables and rates of relapse, after accounting for the influence of the other variables, see Figure 1.

Figure 1. Association between individual predictors and probability of relapse

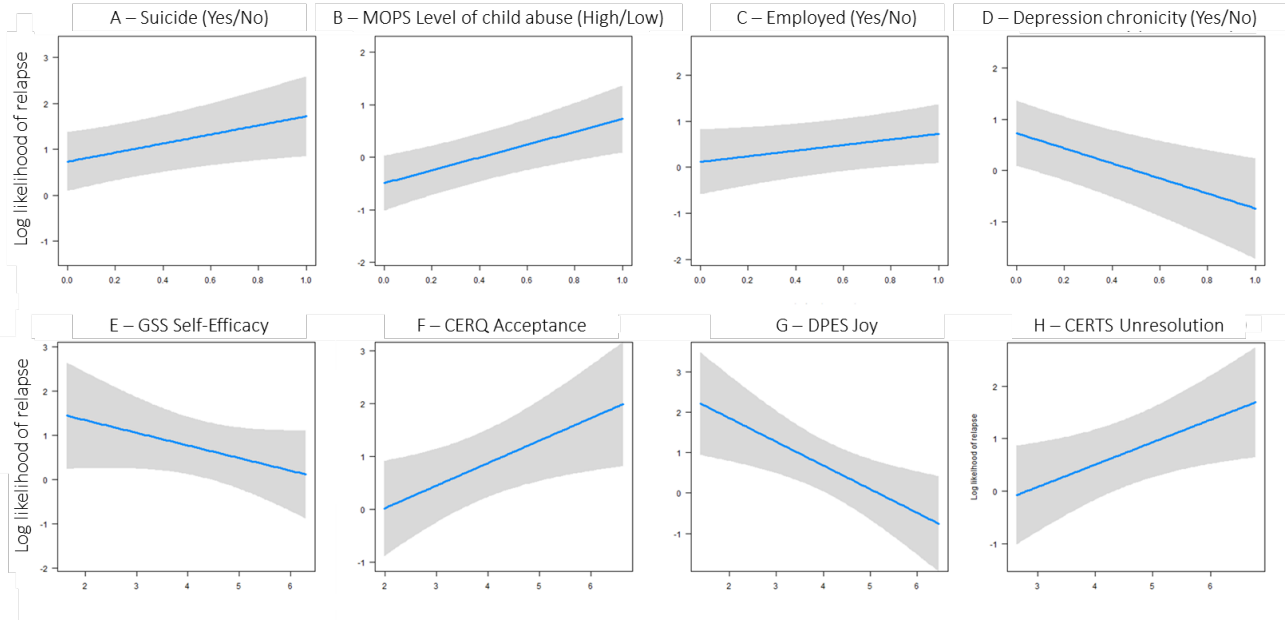

Figure 1. The figure illustrates the association between each predictor in the ADM model and the log likelihood of relapse at 24 months, after accounting for all other variables included in the ADM model. The panels A – D plot the log likelihood of relapse over 24 months for binary variables, with 0 = No/Low versus 1 = Yes/High. Panels E – H plot the association of continuous variables with the log likelihood of relapse. All continuous variables were scaled. MOPS = Measure of Parenting Style; GSS = General Self-Efficacy Scale; CERQ=Cognitive Emotion Regulation Questionnaire; DPES=Dispositional Positive Emotions Scale; CERTS-Cambridge-Exeter Repetitive Thought Scale.

Table 2. Predictor weightings for the final ADM prognostic model

Predictor	*β*	*SE*	*z*	*p*
Intercept	−1.10	1.42	−0.78	.44
Previous suicide attempt (Yes/No)	0.99	0.42	2.33	.02
MOPS Level of child abuse (Low/High)	1.22	0.36	3.40	.001

Employment status (Yes/No)	0.61	0.36	1.68	.09
Previous depressive episode chronicity	−1.47	0.46	−3.17	.002
GSS Self-Efficacy	−0.28	0.19	−1.46	.14
CERQ Acceptance	0.43	0.18	2.38	.02
DPES Dispositional joy	−0.59	0.21	−2.78	.005
CERTS Unresolution	0.43	0.19	2.89	.02

Table 2. The table reports regression coefficients for all predictors in the ADM prognostic model built using logistic regression. In the model, all continuous variables entered were scaled and binary variables were set to 0 and 1. MOPS = Measure of Parenting Style; GSS = General Self-Efficacy Scale; CERQ=Cognitive Emotion Regulation Questionnaire; DPES=Dispositional Positive Emotions Scale; CERTS-Cambridge-Exeter Repetitive Thought Scale.

We then used observed depressive relapse (yes/no) over 24 months to evaluate the factual predictions in the ADM model that were made without the use of each patient's own data. The average AUC estimate of model fit for the ADM model was 0.66 (Figure 2A) suggesting sufficient predictive signal in the variables selected by ENRR. In contrast, the ADM-specific model based only on baseline depression severity on the HAMD in ADM was essentially at chance (AUC = 0.51). The DeLong test for two correlated AUCs showed a significant difference in these AUCs for the HAMD vs. ADM models, $z = 3.11$, $p = .002$.

Figure 2. Probability of relapse in the ADM model

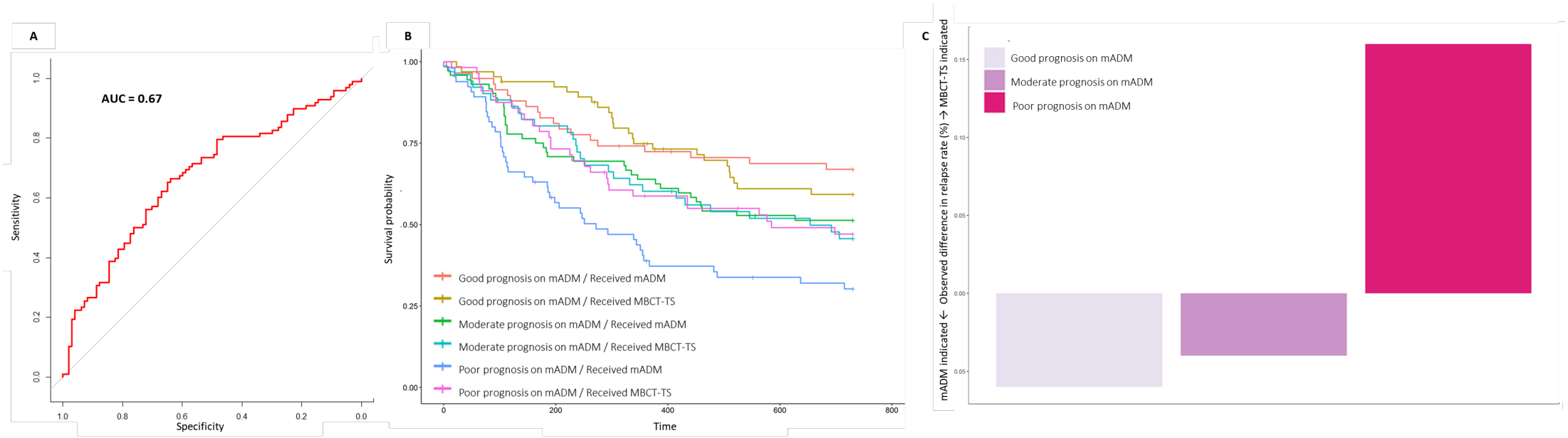

Figure 2. **Panel A** demonstrates the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) delineates the relative sensitivity (true positive rate) and specificity (false positive rate) of the ADM model including the eight variables selected by the elastic net predicting rates of depressive relapse. The AUC (red line) is plotted against the straight grey line, which represents the threshold at which the model has no predictive utility. The grey line delineates the likelihood of someone above and below that threshold on the prognostic index has an equal likelihood of relapse. That is, the larger (further away from the grey line) the AUC the greater a model's predictive utility. **Panel B** plots the predicted survival curves for time (measured in days) to depressive relapse over the two year follow-up period for each ADM-prognostic group (poor, moderate, good) as a function of the treatment they received (MBCT or ADM). Relapse rates for each bin separately were: Good ADM-prognosis, Received ADM = 37.14%, Good ADM-prognosis, Received MBCT = 33.96%, Moderate ADM-prognosis, Received ADM = 44.23%, Moderate ADM-prognosis, Received MBCT = 53.52%, Poor

ADM-prognosis, Received ADM = 67.12%, Poor ADM-prognosis, Received MBCT = 52.08%. **Panel C** summarizes the computed differences in relapse rates over the two year follow up as a function of the treatment received for each ADM-prognostic group.

Prognostic utility

We first verified that the outcome data for our analysis sample were comparable to that of the total PREVENT sample (Kuyken et al., 2015). As in the full sample, survival times (z = −1.02; P = .31, hazard ratio [MBCT relative to ADM] = 0.86; 95%CI, 0.64 to 1.15) and relapse rates (MBCT = 47.1%, ADM = 50.3%) during the 24-month follow-up period in our analysis sample did not significantly differ between the two treatment conditions. (For additional details on all survival analyses in this section, see supplement). In the survival analysis examining time-to-relapse with main effects for treatment and continuous ADM prognosis, there was a significant main effect of continuous ADM prognosis (z = 4.615; P < .001). We next compared observed outcomes across the two treatment conditions for individuals according to their ADM-prognosis (i.e., good, moderate, poor; Figure 2B).

The predicted survival curves did not differ across treatments for those with good ADM prognoses (hazard ratio reflecting increased risk of relapse for those in MBCT vs. ADM = 1.34; 95%CI, 0.73 to 2.45; P = .35). The same was true for those with moderate ADM prognoses (hazard ratio = 1.19; 95%CI, 0.73 to 1.96; P = .48). In contrast, those with poor ADM prognoses were predicted to have significantly reduced relapse risk (hazard ratio = 0.52; 95%CI, 0.32 to 0.84; P = .008) if switching to MBCT instead of continuing with ADM.

When comparing rates of those who had actually relapsed/recurred by the end of the two-year follow-up period, the same pattern emerged. There was a significant main effect of ADM-prognosis on observed relapse rates, X^2 (2) = 16.16, P < .001. As expected, the individuals with a good ADM prognosis showed the lowest rates of relapse (35%), the group with moderate prognosis showed an intermediate relapse rate (51%) and the group with the poor prognosis showed the highest rate of relapse (60%). Relapse rates were low for those with good ADM prognoses regardless of which treatment they received (ADM = 31%, MBCT = 38%). Relapse rates did not differ significantly as a function of treatment assignment for this group (X^2 (1) =

0.45, $P = .50$), or for those with moderate ADM prognoses: X^2 (1) = 0.71, $P = .40$. However, for individuals with poor ADM prognoses, relapse rates were significantly worse for those who received ADM (70%) versus those who received MBCT (48%): X^2 (1) = 4.86, $P = .03$. Finally, results from the sensitivity analyses which repeated the above analyses in all participants with sufficient data (i.e., including those MBCT participants who had been excluded for not having attended at least 4 sessions of MBCT (n=25)) aligned with the results from the primary analysis sample (Supplemental Materials).

DISCUSSION

Clinical depression is a heterogenous condition, which often runs a relapsing and remitting course across the lifespan and where no single treatment has been shown to be effective for all patients (Fried, 2017; Fried & Nesse, 2015). A precision medicine approach to depressive relapse prevention has potential utility in facilitating clinical choices between maintenance pharmacotherapy regimens and preventive psychosocial interventions such as MBCT.

We described a prognostic model that was developed using baseline data (demographic, clinical and readily available psychological measures) from individuals randomized to receive maintenance antidepressant following a successful course of acute treatment with ADM (see Supplementary Materials for further discussion of the model elements). This model, which predicts depressive relapse across a 24 month follow-up period, performed comparably to algorithms predicting acute remission response to antidepressants (Chekroud et al., 2016, 2017; Iniesta et al., 2016). We then generated ADM prognoses for the entire RCT sample (including those randomized to receive MBCT) to investigate whether the information from the ADM predictions might be helpful in deciding between continuing antidepressants or switching to preventative psychotherapy (MBCT). We observed a large difference in relapse rates for patients with poor ADM prognoses: 70% relapse in ADM vs 48% relapse in MBCT. In other words,

patients with a poor prognosis on ADM do not seem to simply be clinical non-responders but, rather, they may be individuals for whom MBCT putatively represents a clinically beneficial alternative. The survival model's estimate of a 48% reduction in risk of relapse across the 24-month follow-up period (hazard ratio = .52) for patients with poor ADM prognoses who received MBCT versus ADM, if it were to be replicated, would suggest that such patients should pursue MBCT.

Given the low relapse rates and lack of difference between treatments for those with good ADM prognoses (31% ADM vs. 38% MBCT), such patients could be encouraged to select which relapse prevention strategy to pursue based on other factors. Clinically, our data indicate that treatment selection for depressive relapse prevention in individuals with recurrent depression who have a moderate to good ADM prognoses could be guided by factors such as patient preference, cost, and resource availability. For individuals with a poor prognosis on ADM, however, our data indicate that MBCT alongside tapering or cessation of medication to prevent relapse potentially confers a better clinical outcome and should be offered as an alternative to ADM. Recent systematic reviews and individual-participant meta-analyses suggest that combination relapse prevention, in which both medication continuation and preventative psychotherapy are provided, is superior to monotherapy, and thus should also be considered for patients at higher risk for relapse (add cite for recent Pim paper and Claudi's paper).

Our study has a number of potential limitations. With the present data we are unable to disaggregate the effects of MBCT from the tapering or discontinuation of ADM, as they were both part of the MBCT protocol. We are also unable to comment on whether the effects are specific to MBCT or whether any effective alternative psychosocial intervention would offer potentially similar benefits for individuals with a poor prognosis on ADM.

The utility of any model depends on its ability to generalise. The present algorithm was subject to internal validation during variable selection and model building. The imputation of

missing baseline data was not included in the cross-validation, but given the low number of missing datapoints, we were not concerned about this as a potential source of significant bias. Recent work suggests that penalization and shrinkage methods may not provide as much protection as is assumed, and that such methods (including ENRR) can produce unreliable clinical prediction models when sample sizes are small (Riley et al., 2021). Despite the internal cross-validation, we were not able to externally validate the model on a wholly independent sample, as comparable sufficiently-large trials evaluating the same preventive interventions, with the range of same or similar baseline measures, are simply not currently available. This reflects the current state of precision medicine research (Cohen & DeRubeis, 2018), in which predictive models are too rarely subjected to proper external validation (Salazar de Pablo et al., 2020). Further external validation, when suitable data become available, will be an important next step prior to the translation of the current findings into firm treatment recommendations.

Ideally, both the ADM and MBCT models would have been sufficiently robust to actively compare the two predictive indices to elucidate what works best for whom. However, our computed MBCT model did not perform above chance and was no better than a prediction model built solely on baseline depression severity scores. This lack of robust prediction within the MBCT model accords with the replicated finding that very few demographic, clinical or psychological variables, over and above baseline symptom severity, appear to predict outcome to MBCT (Kuyken et al., 2016; Kuyken, Watkins, et al., 2010), testifying to the intervention's broad suitability. Secondly, in the present study, MBCT was combined with support for medication tapering or discontinuation and it may be that the mixture of these two different intervention components (and possible associated effects of medication withdrawal) obscured any clear relations in the MBCT arm with the predictor variables included here.

The current findings represent a significant first step in the application of precision medicine to inform patient and clinician choice around optimal interventions for depressive

relapse prevention. Additional work is needed to further validate the model reported here in wholly independent, yet-to-be-collected, large samples. The eventual success of this and similar personalized medicine approaches to mental health care will depend on the acquisition and dissemination of large-scale clinical datasets which will allow for the development and validation of predictive models (Chekroud et al., 2017, 2021). The utility of these models must then be evaluated in prospective clinical trials (Delgadillo & Lutz, 2020), which have begun to emerge with promising results (e.g., Lutz et al., 2021; Delgadillo et al., under review).

References

Altman, D. G., & Bland, J. M. (1994). Statistics Notes: Quartiles, quintiles, centiles, and other quantiles. *British Medical Journal*, *309*(6960), 996-996. https://doi.org/10.1136/bmj.309.6960.996

Austin, P. C., & Steyerberg, E. W. (2012). Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. *BMC Med Res Methodol*, *12*, 82. https://doi.org/10.1186/1471-2288-12-82

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Cooperation.

Bet, P. M., Hugtenburg, J. G., Penninx, B. W. J. H., & Hoogendijk, W. J. G. (2013). Side effects of antidepressants during long-term use in a naturalistic setting. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, *23*(11), 1443–1451. https://doi.org/10.1016/j.euroneuro.2013.05.001

Bonnett, L. J., Snell, K. I., Collins, G. S., & Riley, R. D. (2019). Guide to presenting clinical prediction models for use in clinical settings. *British Medical Journal*, *365*, 1737. https://doi.org/10.1136/bmj.l737

Bosman, R. C., Waumans, R. C., Jacobs, G. E., Oude Voshaar, R. C., Muntingh, A. D. T., Batelaan, N. M., & van Balkom, A. (2018). Failure to Respond after Reinstatement of Antidepressant Medication: A Systematic Review. *Psychother Psychosom*, *87*(5), 268-275. https://doi.org/10.1159/000491550

Breedvelt, J. J. F., Brouwer, M. E., Harrer, M., Semkovska, M., Ebert, D. D., Cuijpers, P., & Bockting, C. L. H. (2020). Psychological interventions as an alternative and add-on to antidepressant medication to prevent depressive relapse: systematic review and meta-analysis. *Br J Psychiatry*, 1-8. https://doi.org/10.1192/bjp.2020.198

Buckman JEJ, Cohen ZD, O'Driscoll C, Fried EI, Saunders R, Ambler G, DeRubeis RJ, Gilbody S, Hollon SD, Kendrick T, Watkins E, Eley TC, Peel AJ, Rayner C, Kessler D, Wiles N, Lewis G, & Pilling S. (2021). Predicting prognosis for adults with depression using individual symptom data: a comparison of modelling approaches. *Psychological Medicine,* 1-11. https://doi.org/10.1017/S0033291721001616

Cartwright, C., Gibson, K., Read, J., Cowan, O., & Dehar, T. (2016). Long-term antidepressant use: Patient perspectives of benefits and adverse effects. *Patient Preference and Adherence*, *10*, 1401–1407. https://doi.org/10.2147/PPA.S110632

Chekroud, A. M., Gueorguieva, R., Krumholz, H. M., Trivedi, M. H., Krystal, J. H., & McCarthy, G. (2017). Reevaluating the Efficacy and Predictability of Antidepressant Treatments: A Symptom Clustering Approach. *JAMA Psychiatry*, *74*(4), 370–378. https://doi.org/10.1001/jamapsychiatry.2017.0025

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

Chekroud A, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, Cohen ZD, Belgrave D, DeRubeis R, Iniesta R, Dwyer D, & Choi K. The promise of machine learning in predicting treatment outcomes in psychiatry. (*In Press at World Psychiatry*)

Cleare, A., Pariante, C. M., Young, A. H., Anderson, I. M., Christmas, D., Cowen, P. J., Dickens, C., Ferrier, I. N., Geddes, J., Gilbody, S., Haddad, P. M., Katona, C., Lewis, G., Malizia, A., McAllister-Williams, R. H., Ramchandani, P., Scott, J., Taylor, D., & Uher, R. (2015). Evidence-based guidelines for treating depressive disorders with antidepressants: A revision of the 2008 British Association for Psychopharmacology guidelines. *Journal of Psychopharmacology*, *29*(5), 459–525. https://doi.org/10.1177/0269881115581093

Cohen, Z. D., Ashar, Y. K., & Derubies, R. J. (2019). Moving Beyond "One Size Fits All.". in *Evidence-Based Practice in Action: Bridging Clinical Science and Intervention*, Sona Dimidjian (Editor). pages 111-132. Guilford Publications.

Cohen Z. D., Delgadillo J, DeRubeis R. J. (2021) Personalized treatment approaches. In: Barkham M., Lutz W., Castonguay L.G. eds. Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, 7th edn. New Jersey: Wiley Blackwell. ISBN: 978-1-119-53658-1

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

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2019). A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive–behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychotherapy Research*, *0*(0), 1–14. https://doi.org/10.1080/10503307.2018.1563312

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

Cuijpers, P., Hollon, S. D., van Straten, A., Bockting, C., Berking, M., & Andersson, G. (2013). Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. *BMJ Open*, *3*(4), e002542. https://doi.org/10.1136/bmjopen-2012-002542

De Zwart, P., Jeronimus, B., & De Jonge, P. (2019). Empirical evidence for definitions of episode, remission, recovery, relapse and recurrence in depression: A systematic review. *Epidemiology and Psychiatric Sciences, 28*(5), 544-562. doi:10.1017/S2045796018000227

Deisenhofer, A.-K., Delgadillo, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety*, *35*(6), 541–550. https://doi.org/10.1002/da.22755

Delgadillo, J., & Lutz, W. (2020). A Development Pathway Towards Precision Mental Health Care. *JAMA Psychiatry*, *77*(9), 889-890. https://doi.org/10.1001/jamapsychiatry.2020.1048

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., Cohen, Z. D., DeRubeis, R. J., and Barkham, M. (*under review*). StratCare: A pragmatic, multi-site, single-blind, cluster randomised controlled trial of stratified care for depression.

DeRubeis, R. J., Zajecka, J., Shelton, R. C., Amsterdam, J. D., Fawcett, J., Xu, C., Young, P. R., Gallop, R., & Hollon, S. D. (2019). Prevention of Recurrence After Recovery From a Major Depressive Episode With Antidepressant Medication Alone or in Combination With Cognitive Behavioral Therapy: Phase 2 of a 2-Phase Randomized Clinical Trial. *JAMA Psychiatry, 77*(3), 237-245. https://doi.org/10.1001/jamapsychiatry.2019.3900

Dorow, M., Löbner, M., Pabst, A., Stein, J., & Riedel-Heller, S. G. (2018). Preferences for depression treatment including internet-based interventions: Results from a large sample of primary care patients. *Frontiers in Psychiatry*, *9*. https://doi.org/10.3389/fpsyt.2018.00181

Fornaro, M., Anastasia, A., Novello, S., Fusco, A., Pariano, R., De Berardis, D., Solmi, M., Veronese, N., Stubbs, B., & Vieta, E. (2019). The emergence of loss of efficacy during antidepressant drug treatment for major depressive disorder: an integrative review of evidence, mechanisms, and clinical implications. *Pharmacological research*, *139*, 494-502.

Frank E, Prien RF, Jarrett RB, *et al*. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. remission, recovery, relapse, and recurrence. Arch Gen Psychiatry1991;**48**:851.doi:10.1001/archpsyc.1991.01810330075011

Fried, E. I. (2017). Moving forward: How depression heterogeneity hinders progress in treatment and research. *Expert Review of Neurotherapeutics*, *17*(5), 423–425. https://doi.org/10.1080/14737175.2017.1307737

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders*, *172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Gelenberg, A. J., Freeman, M. P., Markowitz, J. C., Rosenbaum, J. F., Thase, M. E., Trivedi, M. H., & Van Rhoads, R. S. (2010). American Psychiatric Association practice guidelines for the treatment of patients with major depressive disorder. *American Journal of Psychiatry*, *167*(Suppl. 10), 9–118.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, *6*(4), 278–296.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.

Hong, S., & Lynn, H. S. (2020). Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. *BMC Med Res Methodol*, *20*(1), 199. https://doi.org/10.1186/s12874-020-01080-1

Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M. Z., Souery, D., Stahl, D., Dobson, R., Aitchison, K. J., Farmer, A., Lewis, C. M.,

McGuffin, P., & Uher, R. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric Research*, *78*, 94–102. https://doi.org/10.1016/j.jpsychires.2016.03.016

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Linear model selection and regularization. In An introduction to statistical learning (pp. 203-264). Springer, New York, NY.

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, *34*(1), 119–138. https://doi.org/10.1146/annurev-publhealth-031912-114409

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

Kim, T. T., Dufour, S., Xu, C., Cohen, Z. D., Sylvia, L., Deckersbach, T., DeRubeis, R. J., & Nierenberg, A. A. (2019). Predictive modeling for response to lithium and quetiapine in bipolar disorder. *Bipolar Disorders*, *21*(5), 428-436. https://doi.org/10.1111/bdi.12752

Kinrys, G., Gold, A. K., Pisano, V. D., Freeman, M. P., Papakostas, G. I., Mischoulon, D., Nierenberg, A. A., & Fava, M. (2019). Tachyphylaxis in major depressive disorder: A review of the current state of research. *Journal of Affective Disorders*, *245*, 488-497. https://doi.org/10.1016/j.jad.2018.10.357

Kuyken, W., Byford, S., Byng, R., Dalgleish, T., Lewis, G., Taylor, R., Watkins, E. R., Hayes, R., Lanham, P., Kessler, D., Morant, N., & Evans, A. (2010). Study protocol for a randomized controlled trial comparing mindfulness-based cognitive therapy with

maintenance anti-depressant treatment in the prevention of depressive relapse: The PREVENT trial. *Trials*, *11*, 99. https://doi.org/10.1186/1745-6215-11-99

Kuyken, W., Byford, S., Byng, R., Dalgleish, T., Lewis, G., Taylor, R., Watkins, E. R., Hayes, R., Lanham, P., Kessler, D., Morant, N., & Evans, A. (2014). Update to the study protocol for a randomized controlled trial comparing mindfulness-based cognitive therapy with maintenance anti-depressant treatment depressive relapse: The PREVENT trial. *Trials*, *15*, 217. https://doi.org/10.1186/1745-6215-15-217

Kuyken, W., Hayes, R., Barrett, B., Byng, R., Dalgleish, T., Kessler, D., Lewis, G., Watkins, E., Brejcha, C., Cardy, J., Causley, A., Cowderoy, S., Evans, A., Gradinger, F., Kaur, S., Lanham, P., Morant, N., Richards, J., Shah, P., … Byford, S. (2015). Effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse or recurrence (PREVENT): A randomised controlled trial. *Lancet*, *386*(9988), 63–73. https://doi.org/10.1016/S0140-6736(14)62222-4

Kuyken, W., Hayes, R., Barrett, B., Byng, R., Dalgleish, T., Kessler, D., Lewis, G., Watkins, E., Morant, N., Taylor, R. S., & Byford, S. (2015). The effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse: Results of a randomised controlled trial (the PREVENT study). *Health Technology Assessment*, *19*(73), 1–124. https://doi.org/10.3310/hta19730

Kuyken, W., Warren, F. C., Taylor, R. S., Whalley, B., Crane, C., Bondolfi, G., Hayes, R., Huijbers, M., Ma, H., Schweizer, S., Segal, Z., Speckens, A., Teasdale, J. D., van Heeringen, K., Williams, M., Byford, S., Byng, R., & Dalgleish, T. (2016). Efficacy of Mindfulness-Based Cognitive Therapy in Prevention of Depressive Relapse: An

Individual Patient Data Meta-analysis From Randomized Trials. *JAMA Psychiatry*, *73*(6), 565–574. https://doi.org/10.1001/jamapsychiatry.2016.0076

Kuyken, W., Watkins, E. R., Holden, E., White, K., Taylor, R. S., Byford, S., Evans, A., Radford, S., Teasdale, J. D., & Dalgleish, T. (2010). How does mindfulness-based cognitive therapy work? *Behaviour Research and Therapy*, *48*(11), 1105–1112. https://doi.org/10.1016/j.brat.2010.08.003

Kwan, B. M., Dimidjian, S., & Rizvi, S. L. (2010). Treatment preference, engagement, and clinical improvement in pharmacotherapy versus psychotherapy for depression. *Behaviour Research and Therapy*, *48*(8), 799–804. https://doi.org/10.1016/j.brat.2010.04.003

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*, 2167702618815466. https://doi.org/10.1177/2167702618815466

Lutz, W., Deisenhofer, A.-K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2021). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology*, *in press*, https://doi.org/10.1037/ccp0000642.

McCartney, M., Nevitt, S., Lloyd, A., Hill, R., White, R., & Duarte, R. (2021). Mindfulness-based cognitive therapy for prevention and time to depressive relapse: Systematic review and network meta-analysis. *Acta Psychiatr Scand*, *143*(1), 6-21. https://doi.org/10.1111/acps.13242

McCrea, R. L., Sammon, C. J., Nazareth, I., & Petersen, I. (2016). Initiation and duration of selective serotonin reuptake inhibitor prescribing over time: UK cohort study. *The British*

Journal of Psychiatry: The Journal of Mental Science, *209*(5), 421–426. https://doi.org/10.1192/bjp.bp.115.166975

McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013). Patient preference for psychological vs. pharmacological treatment of psychiatric disorders: A meta-analytic review. *Journal of Clinical Psychiatry*, *74*(6), 595–602. https://doi.org/10.4088/JCP.12r07757

Mojtabai, R., & Olfson, M. (2014). National trends in long-term use of antidepressant medications: Results from the U.S. National Health and Nutrition Examination Survey. *The Journal of Clinical Psychiatry*, *75*(2), 169–177. https://doi.org/10.4088/JCP.13m08443

NICE. (2009). *Depression—Guidance and guidelines—NICE*. https://www.nice.org.uk/guidance/cg90/chapter/1-guidance

OECD. (2013). *Health at a glance 2013: OECD Indicators*. OECD Publisher. https://doi.org/10.1787/health_glance-2013-en

Parker, G., Roussos, J., Hadzi-Pavlovic, D., Mitchell, P., Wilhelm, K., & Austin, M. P. (1997). The development of a refined measure of dysfunctional parenting and assessment of its relevance in patients with affective disorders. *Psychological Medicine*, *27*(5), 1193–1203.

Patel, V., Chisholm, D., Parikh, R., Charlson, F. J., Degenhardt, L., Dua, T., Ferrari, A. J., Hyman, S., Laxminarayan, R., Levin, C., Lund, C., Medina Mora, M. E., Petersen, I., Scott, J., Shidhaye, R., Vijayakumar, L., Thornicroft, G., & Whiteford, H. (2016). Addressing the burden of mental, neurological, and substance use disorders: Key messages from Disease Control Priorities , 3rd edition. *Lancet*, *387*(10028), 1672–1685. https://doi.org/10.1016/S0140-6736(15)00390-6

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. M., Huang, Y., Jordans, M. J.

D., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., … UnÜtzer, Jü. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, *392*(10157), 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X

Pearson, R., Pisner, D., Meyer, B., Shumake, J., & Beevers, C. G. (2018). A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression. *Psychological Medicine*, *49*(14), 2330-2341. https://doi.org/https://doi.org/10.1017/S003329171800315X

Piek, E., van der Meer, K., & Nolen, W. A. (2010). Guideline recommendations for long-term treatment of depression with antidepressants in primary care—A critical review. *The European Journal of General Practice*, *16*(2), 106–112. https://doi.org/10.3109/13814781003692463

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Riley, R. D., Snell, K. I. E., Martin, G. P., Whittle, R., Archer, L., Sperrin, M., & Collins, G. S. (2021). Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. *J Clin Epidemiol*, *132*, 88-96. https://doi.org/10.1016/j.jclinepi.2020.12.005

Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., Baldwin, H., Danese, A., Fazel, S., Steyerberg, E. W., Stahl, D., & Fusar-Poli, P. (2020). Implementing Precision Psychiatry: A Systematic Review of Individualized Prediction Models for Clinical Practice. *Schizophr Bull*. https://doi.org/10.1093/schbul/sbaa120

Schweizer, S., Peeters, F., Huibers, M., Roelofs, J., van Os, J., & Arntz, A. (2010). Does illness attribution affect treatment assignment in depression? *Clinical Psychology & Psychotherapy*, *17*(5), 418–426.

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology*, *179*(6), 764–774. https://doi.org/10.1093/aje/kwt312

Shay, L. A., & Lafata, J. E. (2015). Where Is the Evidence? A Systematic Review of Shared Decision Making and Patient Outcomes. *Medical Decision Making*, *35*(1), 114–131. https://doi.org/10.1177/0272989X14551638

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports*, *51*(2), 663–671.

Stacey, D., Legare, F., Lewis, K., Barry, M. J., Bennett, C. L., Eden, K. B., Holmes-Rovner, M., Llewellyn-Thomas, H., Lyddiatt, A., Thomson, R., & Trevena, L. (2017). Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews*, *4*(4), CD001431. https://doi.org/10.1002/14651858.CD001431.pub5

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

Waldmann, P., Mészáros, G., Gredler, B., Fuerst, C., & Sölkner, J. (2013). Evaluation of the lasso and the elastic net in genome-wide association studies. *Frontiers in Genetics, 4,* 1–11. http://dx.doi.org/10.3389/fgene .2013.00270

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., & Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, *3*(8), e002847. https://doi.org/10.1136/bmjopen-2013-002847

Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: a comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology*, 88(1), 25–38. https://doi.org/10.1037/ccp0000451

Weston, W. W. (2001). Informed and shared decision-making: The crux of patient-centered care. *Canadian Medical Association Journal*, *165*(4), 438–439.

Williams, J. B., Kobak, K. A., Bech, P., Engelhardt, N., Evans, K., Lipsitz, J., Olin, J., Pearson, J., & Kalali, A. (2008). The GRID-HAMD: standardization of the Hamilton Depression Rating Scale. *International Clinical Psychopharmacology*, *23*(3), 120-129. https://doi.org/10.1097/YIC.0b013e3282f948f5

Windle, E., Tee, H., Sabitova, A., Jovanovic, N., Priebe, S., & Carr, C. (2019). Association of patient treatment preference with dropout and clinical outcomes in adult psychosocial mental health interventions: A systematic review and meta-analysis. *JAMA Psychiatry, 77*(3), 294-302. https://doi.org/10.1001/jamapsychiatry.2019.3750

Winston, K., Grendarova, P., & Rabi, D. (2018). Video-based patient decision aids: A scoping review. *Patient Education and Counseling*, *101*(4), 558-578. https://doi.org/10.1016/j.pec.2017.10.009

Wiltsey Stirman, S., Cohen, Z. D., Lunney, C. A., DeRubeis, R. J., Wiley, J. F., & Schnurr, P. P. (2021). A Personalized Index to Inform Selection of a Trauma-Focused or Non-Trauma-Focused Treatment for PTSD. *Behaviour Research and Therapy*. https://doi.org/10.1016/j.brat.2021.103872

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301-320.