



# A prognostic index for long-term outcome after successful acute phase cognitive therapy and interpersonal psychotherapy for major depressive disorder

Suzanne C. van Bronswijk MD<sup>1,2</sup>  | Lotte H.J.M. Lemmens PhD<sup>3</sup> |  
 John R. Keefe MA<sup>4,5</sup>  | Marcus J.H. Huibers PhD<sup>4,6</sup> | Robert J. DeRubeis PhD<sup>4</sup> |  
 Frenk P.M.L. Peeters MD, PhD<sup>1,2</sup>

<sup>1</sup>Department of Psychiatry and Psychology, Maastricht University Medical Center, Maastricht, the Netherlands

<sup>2</sup>School for Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands

<sup>3</sup>Department of Clinical Psychological Science, Maastricht University, Maastricht, the Netherlands

<sup>4</sup>Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States

<sup>5</sup>Department of Psychiatry, Weill Cornell Medical College, New York, United States

<sup>6</sup>Department of Clinical Psychology, VU University Amsterdam, Amsterdam, the Netherlands

## Correspondence

Suzanne van Bronswijk, Department of Psychiatry and Neuropsychology, Faculty of Health, Medicine and Life Sciences, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands.  
 Email: [suzanne.vanbronswijk@maastrichtuniversity.nl](mailto:suzanne.vanbronswijk@maastrichtuniversity.nl)

**Background:** Major depressive disorder (MDD) has a highly recurrent nature. After successful treatment, it is important to identify individuals who are at risk of an unfavorable long-term course. Despite extensive research, there is no consensus yet on the clinically relevant predictors of long-term outcome in MDD, and no prediction models are implemented in clinical practice. The aim of this study was to create a prognostic index (PI) to estimate long-term depression severity after successful and high quality acute treatment for MDD.

**Methods:** Data come from responders to cognitive therapy (CT) and interpersonal psychotherapy (IPT) in a randomized clinical trial ( $n = 85$ ; CT = 45, IPT = 40). Primary outcome was depression severity, assessed with the Beck Depression Inventory II, measured throughout a 17-month follow-up phase. We examined 29 variables as potential predictors, using a model-based recursive partitioning method and bootstrap resampling in conjunction with backwards elimination. The selected predictors were combined into a PI. Individual PI scores were estimated using a cross-validation approach.

**Results:** A total of three post-treatment predictors were identified: depression severity, hopelessness, and self-esteem. Cross-validated PI scores evidenced a strong correlation ( $r = 0.60$ ) with follow-up depression severity.

**Conclusion:** Long-term predictions of MDD are multifactorial, involving a combination of variables that each has a small prognostic effect. If replicated and validated, the PI can be implemented to predict follow-up depression severity for each individual after acute treatment response, and to personalize long-term treatment strategies.

## KEYWORDS

clinical trials, cognitive therapy, depression, empirical supported treatments, interpersonal psychotherapy

## 1 | INTRODUCTION

One of the major challenges in the treatment of major depressive disorder (MDD) is its highly recurrent nature, with estimated recurrence rates of up to 90% (American Psychiatric Association, 2000; Solomon

et al., 2000). Two of the most well-studied and frequently applied psychotherapies for MDD are cognitive therapy (CT) and interpersonal psychotherapy (IPT). Although the majority of the individuals who respond to acute phase CT and IPT maintain their depressive symptom reduction, approximately one third relapse within a 17-month

follow-up period (Lemmens et al., 2018). For accurate clinical decision making, it is important to distinguish individuals who are at risk of an unfavorable long-term course from those who will likely maintain their improvement after terminating successful treatment. Additional monitoring and maintenance treatment should be considered for at risk individuals to prevent relapse and recurrence. Maintenance treatment options such as long-term pharmacotherapy and psychological interventions have been proven effective following acute phase treatment (Biesheuvel-Leliefeld et al., 2015; Hansen et al., 2008). However, identifying at risk individuals remains challenging, and current clinical decision-making results in both under- and overtreatment of recovered depressed individuals (Lorenzo-Luaces, 2015; Monroe & Harkness, 2011).

In the field of medicine, it is increasingly common to improve and support clinical decision making using prediction algorithms. With these algorithms, several relevant variables, also referred to as predictors, are combined, resulting in a quantified estimate of the individual's prognosis. These estimates, also referred to as scores on a *prognostic index* (PI), can be used to determine the level of care that is needed in the future (Lorenzo-Luaces, DeRubeis, van Straten, & Tiemens, 2017). Well-known examples of this approach are the Framingham risk prediction algorithms for cardiovascular disease (Anderson, Wilson, Odell, & Kannel, 1991) and prediction algorithms for cancer risk (Chen et al., 2006; Park et al., 2009).

In clinical psychology and psychiatry, research on multivariate algorithms to predict acute phase treatment outcomes in MDD is rapidly growing. Among others, recent studies have focused on predicting the optimal treatment for a given individual using a so-called "Personalized Advantage Index Approach" (Cohen & DeRubeis, 2018; DeRubeis et al., 2014; Huibers et al., 2015). In addition, personalized predictions to determine the optimal level of treatment intensity have been studied (Delgadoillo, Moreea, & Lutz, 2016; Lorenzo-Luaces et al., 2017). So far, only two studies have investigated multivariate prediction algorithms for the long-term course of MDD (van Loo, Aggen, Gardner, & Kendler, 2015; Wang et al., 2014). These studies—both of an epidemiological nature—have developed powerful prediction models, providing valuable insights into the multifactorial nature of depression recurrence. Unfortunately, because of their naturalistic, population-based design, these studies cannot disentangle treatment effects from natural course.

The development of algorithms for the prediction of long-term course of MDD is impeded by several factors. First of all, evidence is inconclusive for several predictors, including stressful life events, a family history of MDD, severity of the depressive episode, and comorbidity (e.g., dysthymia and personality pathology, Burcusa & Iacono, 2007; Hardeveld et al., 2013; Hardeveld, Spijker, De Graaf, Nolen, & Beekman, 2013). Only a few predictors have been repeatedly found to relate to an unfavorable long-term course of MDD, namely the number of previous episodes, residual symptoms, and childhood maltreatment (Berwian, Walter, Seifritz, & Huys, 2017; Hardeveld, Spijker, De Graaf, Nolen, & Beekman, 2010; Kovacs, Obrosky, & George, 2016; Nanni, Uher, & Danese, 2012). Second, several studies have shown that the risk of recurrence of MDD depends on the combination of multiple predictors each having a small

prognostic effect (van Loo et al., 2015; Wang et al., 2014). As a consequence, slightly weaker predictors are dominated by stronger predictors. Variable selection methods that can deal with a large number of "small-effect" variables are therefore warranted. Third, there is little knowledge how to aggregate this predictive information, especially when the directions of multiple variables conflict. For example, one individual could have several factors placing them at risk for recurrence, and other factors predicting their resilience against recurrence. This information must be combined in some way if it is to be useful clinically.

The aim of the current study was to create a multivariate prediction algorithm to estimate long-term depression severity after successful and high quality acute phase psychotherapy for MDD. With this prediction algorithm, individual scores on a prognostic index can be calculated (PI scores). Within the context of a 17 month follow-up of a recent randomized controlled trial (RCT) comparing CT and IPT (Lemmens et al., 2015, 2018), we explored various clinical and demographic predictors and generated a prediction algorithm. We expected to identify predictors that have been found repeatedly in previous research (i.e., number of previous episodes, residual depression, and childhood maltreatment). In addition, we expected that the aggregation of variables would result in a valid prediction algorithm that provides reliable prognostic information for each individual. If replicated, such an algorithm can be used to identify individuals with unfavorable long-term depression severity, which enables selection of appropriate candidates for maintenance therapy.

## 2 | METHOD

### 2.1 | Design and participants

Data come from an RCT into the effectiveness of individual CT and IPT for MDD. A detailed description of the study design, participants, interventions, and acute and long-term main outcomes is provided elsewhere (Lemmens et al., 2011, 2015, 2018). In brief, depressed outpatients were recruited from the mood disorders unit of the Academic Maastricht Outpatient Mental Health Centre (RIAGG Maastricht, the Netherlands). All participants had a primary diagnosis of MDD as confirmed with the Structured Clinical Interview for DSM-IV Axis I disorders (First, Spitzer, Gibbon, & Williams, 1995). Other inclusions criteria were: internet access, an email address, and sufficient knowledge of the Dutch language. Exclusion criteria were bipolar disorder, highly chronic depression (current episode >5 years), an increased acute suicide risk, concomitant pharmacological or psychological treatment, drugs and alcohol abuse/dependence, and an IQ lower than 80. After obtaining written informed consent, a total of 182 participants were randomly allocated to either CT ( $n = 76$ ), IPT ( $n = 75$ ), or a 2-month waiting-list control condition followed by treatment of choice ( $n = 31$ ). CT and IPT were carried out following the guidelines by Beck, Rush, Shaw, and Emery (1979) for CT, and the guidelines by Klerman, Weissman, Rounsaville, and Chevron (1984) for IPT. Participants received 16–20 sessions ( $M = 17$  sessions,  $SD = 2.9$ ), which were planned weekly and were allowed to be scheduled less frequent

toward the end of treatment. The quality of therapy was rated as “(very) good” to “excellent” by independent assessors in both CT and IPT. The study was approved by the Medical Ethics Committee of Maastricht University Medical Center, and was registered at isrctn.com (identifier: ISRCTN 67561918). Based on our aims, we limited the sample for the current study to 17-month follow-up data from treatment responders to CT and IPT ( $n = 85$ ; CT = 45, IPT = 40). Treatment responders were identified as individuals with either: (1) a post-treatment Beck Depression Inventory II (BDI-II) score lower than 10; or (2) an overall change of at least 9 BDI-II points and a post-treatment BDI-II score lower than 20 (Lemmens et al., 2018). Throughout the follow-up period, individuals were free to pursue additional treatment for MDD. Additional treatment included psychological support (one or more sessions with a general practitioner or a mental health care professional) or the use of antidepressant medication (ADM).

## 2.2 | Measures

### 2.2.1 | Primary outcome

Primary outcome was depression severity assessed with the Beck Depression Inventory, second edition (Beck, Steer, & Brown, 1996; Lemmens et al., 2018). The BDI-II was measured at post-treatment (month 7), monthly for a 5-month period (at months 8, 9, 10, 11, 12), and at the end of the follow-up phase (month 24). BDI-II scores during this 17 month period were aggregated into an Area under the Curve (AUC) to obtain an overall measure for depression severity across the follow-up phase.

### 2.2.2 | Variables

Based on previous work (Fournier et al., 2009; Huibers et al., 2015), we examined 69 variables from six domains as potential predictors: (1) depression variables, (2) demographics, (3) psychological distress, (4) general functioning, (5) psychological processes, and (6) life and family history. We removed redundant variables (i.e., variables that overlapped in content, e.g. subscales of one measurement instrument), by examining the correlations between all variables corrected for attenuation. As shown in Table 1, 29 variables were eventually included in our analyses. The variables were extracted from the following instruments: BDI-II, Beck Hopelessness Scale (BHS, Beck & Steer, 1988), Brief Symptom Inventory (BSI, Derogatis & Melisaratos, 1983), SCID-I, Structured Clinical Interview for DSM-IV Axis I (SCID-I, First, Gibbon, Spitzer, Williams, & Benjamin, 1997), Work and Social Adjustment Scale (WSAS, Mundt, Marks, Shear, & Greist, 2002), RAND-36 (van der Zee & Sanderman, 1993), Dysfunctional Attitudes Scale (DAS, de Graaf, Roelofs, & Huibers, 2009; Weissman & Beck, 1978), Inventory of Interpersonal Problems (IIP, Horowitz, Rosenberg, Baer, Ureno, & Villaseñor, 1988), Self Liking and Self Competence Scale Revised (SLSC-R, Tafarodi & Swann, 2001; Vandromme, Hermans, Spruyt, & Eelen, 2007), Leiden Index of Depression Sensitivity (LEIDS, Van der Does, 2002), Ruminative Response Scale (RRS, Raes, Hermans, & Eelen, 2003), and Attributional Style Questionnaire (ASQ, Cohen, Van den Bout, Kramer, & Van Vliet, 1986; Peterson et al., 1982).

### 2.2.3 | Data analyses

#### Variable description and imputation of missing data

First, we investigated the means and frequencies (where appropriate) of the 29 variables. Second, we examined the number of missing variables and BDI-II outcomes. Third, we imputed missing data with a random forest algorithm (R package “MissForest,” Stekhoven & Bühlmann, 2012). This imputation method is considered to be highly accurate, outperforming other common imputation techniques (Stekhoven & Bühlmann, 2012; Waljee et al., 2013). The following information was included in the imputation model: (1) all nonmissing variables, (2) change during treatment of all nonmissing variables (if available), (3) change from baseline (pre-treatment) of all nonmissing BDI-II outcomes (at 7, 8, 9, 10, 11, 12, and 24 months), and (4) information about the condition (CT/IPT). Finally, we checked the accuracy of our imputation method, by applying it to artificially produced missing data in the subset of individuals with complete data. To estimate the imputation error, imputed values were compared with actual data values by calculating the normalized root mean squared error (NRMSE) for continuous data and the proportion of falsely classified entries (PFC) for categorical data using the “prodNA” function of the R Package “MissForest.”

#### Outcome transformation

In order to get an indication of the “overall” depression severity throughout the complete 17-month follow-up phase, BDI-II scores at 8, 9, 10, 11, 12, and 24 months were aggregated into an AUC using cubic splines to compute integrals. We used the square root of the AUC for further analyses, as the residuals of the AUC outcomes appeared to be non-normally distributed.

#### Variable transformation

Continuous variables were standardized and discrete and categorical variables were centered. For the discrete variables that had a non-normal distribution, log transformations and square root transformations were applied based on visual inspection. For each variable, we describe the applied transformation in Supplemental Methods I.

#### Variable selection

To construct a powerful prediction algorithm, we used two techniques to identify predictors for long-term depression severity from the 29 variables available: a model-based recursive partitioning method followed by bootstrap resampling in conjunction with backwards elimination (R packages “mobForest” and “bootstepAIC,” Garge et al., 2013; Rizopoulos & Rizopoulos, 2009; previous applications of this method: Keefe et al., 2018; Zilcha-Mano et al., 2016). The model-based recursive partitioning technique is based on a random forest algorithm (Garge et al., 2013). In this method, bootstrapped samples are repeatedly split into two subgroups, forming tree-shaped structures. The splits (or nodes) of these trees depend on the variable with the strongest impact on the outcome of a predetermined model. In the current analyses, we applied a regression model with the AUC as the dependent variable. At each node of each tree, a random subset of the potential predictors was applied instead of all variables at once. By doing this, predictors with smaller effects are not eliminated by the

presence of stronger predictors (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008). Parameters were set as follows: 10,000 trees were computed with a minimum  $\alpha$  level of 0.10 for splits and a minimum node size for splitting of 15 individuals. After computing the trees, variables were ranked using a variable-importance plot to identify the predictors with the most accurate predictions. A variable importance score was calculated for each variable to indicate their importance in producing accurate predictions. Importance scores were computed for each variable by subtracting the predictive accuracy of a variable when applying the real values, from the predictive accuracy of a variable when applying randomly permuted values. This difference in accuracy before and after permutation estimates the true predictive impact of a variable. Variables were then ranked based on their importance score and were selected if their importance score was higher than the absolute value of the lowest ranking variable (Garge et al., 2013).

The robustness of these selected variables was then tested by applying backwards elimination on multiple bootstrapped samples (Austin & Tu, 2004). A regression model with the AUC as the dependent variable and the variables selected at the first step as the independent variables was computed. Then, backwards elimination (with  $\alpha = 0.05$ ) of this model was applied to 1,000 bootstrapped samples of the original data. If predictors were selected in at least 60% of the bootstrap samples, they were considered robust and subsequently used for the prediction algorithm. The 60% cut-off has been proven to select variables that construct a parsimonious model with good model fitting (Austin & Tu, 2004). For each predictor, the direction of the estimated regression coefficients was examined by counting the number of times it showed a positive or negative value in the bootstrapped samples.

### Estimating Prognostic Index scores using a 5-fold cross validation

We created a prognostic index by combining all final predictors in the following regression model with the AUC as the dependent variable:  $\text{SQRT}(\text{AUC}_{8-24 \text{ months}}) = \beta_0 + (\beta_1 * \text{Predictor 1}) + (\beta_2 * \text{Predictor 2}) \dots (\beta_n * \text{Predictor } n)$ . A 5-fold cross validation was applied to calculate individual estimates (PI scores). In short, the sample was split into five equal groups and the outcomes of the individuals of each group were predicted using the regression model based on information from the other four groups of the sample (the "training dataset," Picard & Cook, 1984). This technique reduces the risk of overfitting, since the individual whose PI score is calculated is not included in the training dataset of the model used to calculate the score. The PI's were then compared with the actual outcomes. To facilitate interpretation of this comparison, we converted the AUC outcomes to an "average BDI-II score" across the follow-up phase by dividing the AUC by time in months. We calculated the average difference between the actual scores and the predicted scores, and examined the association between these scores using a correlation analysis. We also examined the number of individuals that had an average difference of 5 points or higher on the BDI-II between actual and observed scores, which corresponds to a clinically meaningful difference (Hiroe et al., 2005). In addition, we applied a conditional inference procedure (permutation test) to test the hypothesis of independence between the actual scores and the predicted scores (R package "coin," Hothorn, Hornik, Van De Wiel, & Zeileis, 2006, 2008).

### Additional support during follow-up

We examined the impact of additional treatment (psychological support or antidepressant treatment) on the outcome variable (the AUC) and the PI scores. For this, the AUC and PI scores of individuals with and without additional support were compared using a *t*-test.

### Testing robustness of variable selection and model fitting

Because we performed variable selection and model fitting in the same sample, it is possible that the effect size of how well the PI predicts outcomes will be inflated (Vul, Harris, Winkielman, & Pashler, 2009), even with the statistical precautions taken above. To examine the degree to which the PI's predictive ability may be inflated, we ran a secondary analysis testing the process of both variable selection and model-fitting to predict to a 5-fold held-out sample. This process is less clinically relevant as it produces five potentially different models, but it provides an estimate of how well the entire process selects variables and makes predictions that robustly predict out-of-sample.

## 3 | RESULTS

### 3.1 | Variable description and imputation of missing data

Table 1 describes the means and frequencies (where appropriate) of the 29 variables grouped in five domains (Fournier et al., 2009; Huibers et al., 2015). Of all 29 variables, 8 values were missing (0.003%). A total

**TABLE 1** Sample Description: Depression Variables, Demographics, Psychological Distress Variables, General Functioning Variables, Psychological Processes Variables, and Life and Family History Variables

<b>DOMAIN I: Depression</b>	
Recurrent episodes <sup>a</sup> , <i>n</i> (%)	40 (49.4)
Post-treatment depression severity, BDI-II <sup>b</sup> , mean (S.D.)	7.5 (6.0)
Hopelessness, BHS <sup>b</sup> , mean (S.D.)	4.4 (3.3)
<b>DOMAIN II: Demographics</b>	
Female <sup>a</sup> , <i>n</i> (%)	56 (65.9)
Age <sup>a</sup> , mean (S.D.)	41.6 (11.3)
Partner <sup>a</sup> , <i>n</i> (%)	54 (63.5)
Education level <sup>a</sup>	
Low, <i>n</i> (%)	18 (21.2)
Intermediate, <i>n</i> (%)	48 (56.5)
High, <i>n</i> (%)	19 (22.4)
Active employment <sup>a</sup>	55 (64.7)
Treatment expectancy <sup>a</sup> , mean (S.D.) (0 = not successful; 10 = very successful)	6.8 (1.0)
<b>DOMAIN III: Psychological distress</b>	
General psychological distress, BSI <sup>b</sup> , mean (S.D.)	26.2 (25.2)
Number of comorbid axis I disorders, SCID-I <sup>a</sup> , mean (S.D.)	0.6 (0.7)
Number of comorbid axis II disorders, SCID-II <sup>a</sup> , mean (S.D.)	0.6 (0.8)
Number of comorbid axis II traits, SCID-II <sup>a</sup> , mean (S.D.)	0.5 (0.9)

(Continues)

**TABLE 1** (Continued)

DOMAIN I: Depression	
DOMAIN IV: General functioning	
Social and work functioning, WSAS <sup>b</sup> , mean (S.D.)	11.6 (9.0)
Level of impairment, RAND-36 <sup>b</sup>	
DOMAIN V: Psychological processes	
Dysfunctional Beliefs, DAS <sup>b</sup> , mean (S.D.)	20.5 (6.1)
Interpersonal problems, IIP <sup>b</sup> , mean (S.D.)	55.1 (34.8)
Self Liking and Self Competence, SLSC-R <sup>b</sup> , mean (S.D.)	47.4 (11.4)
Cognitive Reactivity, LEIDS <sup>b</sup> , mean (S.D.)	33.8 (17.0)
Rumination, RRS <sup>b</sup> , mean (S.D.)	37.3 (9.4)
Attributional Style, ASQ <sup>b</sup> , mean (S.D.)	0.7 (1.0)
DOMAIN VI: Life and family history	
Number of life events past year <sup>a</sup> , mean (S.D.)	0.9 (1.3)
Number of childhood trauma events <sup>a</sup> , mean (S.D.)	0.9 (1.3)
Parental: one or both parents <sup>a</sup> , <i>n</i> (%)	
In treatment for illness, <i>n</i> (%)	19 (22.4)
With an anxiety disorder, <i>n</i> (%)	13 (15.3)
With depression, <i>n</i> (%)	34 (40.0)
With alcohol abuse, <i>n</i> (%)	11 (12.9)
With suicidality, <i>n</i> (%)	7 (8.2)

BDI-II, Beck Depression Inventory, second edition; BHS, Beck Hopelessness Scale; BSI, Brief Symptom Inventory; SCID-I, Structured Clinical Interview for DSM-IV Axis I disorders; SCID-II, Structured Clinical Interview for DSM-IV Axis II disorders; WSAS, Work and Social Adjustment Scale; DAS, Dysfunctional Attitudes Scale; IIP, Inventory of Interpersonal Problems; SLSC-R, Self Liking and Self Competence Scale Revised; LEIDS, Leiden Index of Depression Sensitivity; RRS, Ruminative Response Scale; ASQ, Attributional Style Questionnaire.

<sup>a</sup>Measured pretreatment.

<sup>b</sup>Measured post-treatment.

of 79 individuals (93.0%) had no missing values for any of the variables. Of all BDI-II outcomes at 8, 9, 10, 11, 12, and 24 months, 32 outcomes were missing (0.06%). A total of 74 individuals (87.1%) had no missing BDI-II outcomes at any of the time points. The imputation method for the missing variables and outcomes appeared to be accurate, with an estimated normalized root mean squared error (NRMSE) of 0.36 and the estimated proportion of falsely classified entries (PFC) of 0.33 (R package “MissForest,” Stekhoven & Bühlmann, 2012).

### 3.2 | Selection of predictors

We identified 10 potential predictor variables using the model-based recursive partitioning technique. These variables are displayed in Table 2, ranked by their “variable importance.” Of the 10 variables, 3 variables were selected in at least 60% of the bootstrap samples using the backwards elimination technique (Table 3). As shown in this table, for each of the three variables, the regression coefficients across the bootstrap samples appeared to be stable: either all positive or all negative. More specifically, higher post-treatment depression severity (BDI-II) and negative beliefs about the future (Hopelessness, BHS) were associated with higher BDI-II scores across 17 months of follow-up. Higher post-treatment self-esteem (SLSC-R) was associated with lower BDI-II scores during follow-up.

**TABLE 2** Predictors Selected with the Model-Based Recursive Partitioning Technique

Variable Importance Rank (High to Low)	Variable
1	Depression severity, BDI-II
2	Hopelessness, BHS
3	Social and work functioning, WSAS
4	Self Liking and Self Competence, SLSC-R
5	Level of impairment: physical functioning, RAND-36
6	Dysfunctional Beliefs about achievement, DAS
7	General psychological distress, BSI
8	Interpersonal problems, IIP
9	Level of impairment: social functioning, RAND-36
10	Number of life events past year

BDI-II, Beck Depression Inventory, second edition; BHS, Beck Hopelessness Scale; WSAS, Work and Social Adjustment Scale; SLSC-R, Self Liking and Self Competence Scale Revised; DAS, Dysfunctional Attitudes Scale; BSI, Brief Symptom Inventory; IIP, Inventory of Interpersonal Problems.

**TABLE 3** Predictors Selected with Bootstrap Resampling in Conjunction with Backwards Elimination

Variable	Percentage of Bootstrapped Samples This Variable was Selected in	Significant $\beta$	
		Positive	Negative
Depression severity, BDI-II	96.0%	100.0%	0.0%
Self Liking and Self Competence, SLSC-R	81.5%	0.1%	99.9%
Hopelessness, BHS	70.1%	99.7%	0.3%

BDI-II, Beck Depression Inventory, second edition; SLSC-R, Self Liking and Self Competence Scale Revised; BHS, Beck Hopelessness Scale.

### 3.3 | Estimating Prognostic Index scores using a 5-fold cross validation

We combined the three selected predictors into the following regression model to create the PI:  $SQRT(AUC_{8-24\text{ months}}) = \beta_0 + (\beta_1 * BDI-II_{\text{post-treatment}}) + (\beta_2 * BHS_{\text{post-treatment}}) + (\beta_3 * SLSC-R_{\text{post-treatment}})$ . PI scores were estimated using a 5-fold cross validation. As shown in Figure 1, these predicted values (converted to average follow-up BDI-II scores) were compared with the actual average follow-up BDI-II scores. The mean difference between the actual and predicted average BDI-II scores was 0.84 (SD = 5.42) and the correlation between these values was significant and strong (corr. = 0.60,  $p < 0.001$ ). For 21 of the 85 individuals, this difference was clinically significant with 5 points or higher on the BDI-II. Results of the permutation test indicated a significant association between the actual and the predicted scores ( $Z = 5.46$ ,  $p < 0.001$ ). Our secondary analysis examining the estimated PI-outcome correlation that would be obtained without “double-dipping” (i.e., performing variable selection and model fitting in the same sample) did not produce a substantively different correlation than our primary analysis (corr. = 0.62,  $p < 0.001$ ).

The five different regression models are described in Supplemental Results I.

### 3.4 | Additional support

A total of 26 of the 85 responders received additional care: 23 individuals received psychological support (one or more sessions with a general practitioner or a mental health care professional) and 6 were treated with ADM. No differences between the AUC scores of responders with and without additional support were found ( $M = 154.5$  ( $SD = 21.6$ ) vs.  $M = 128.1$  ( $SD = 109.1$ );  $t(79) = -1.01$ ,  $p = 0.31$ ). However, responders with additional care had significant higher overall PI scores as compared to responders without additional support ( $M = 151.0$  ( $SD = 78.8$ ) vs.  $M = 109.7$  ( $SD = 67.9$ );  $t(79) = -2.4192$ ,  $p = 0.0179$ ).

## 4 | DISCUSSION

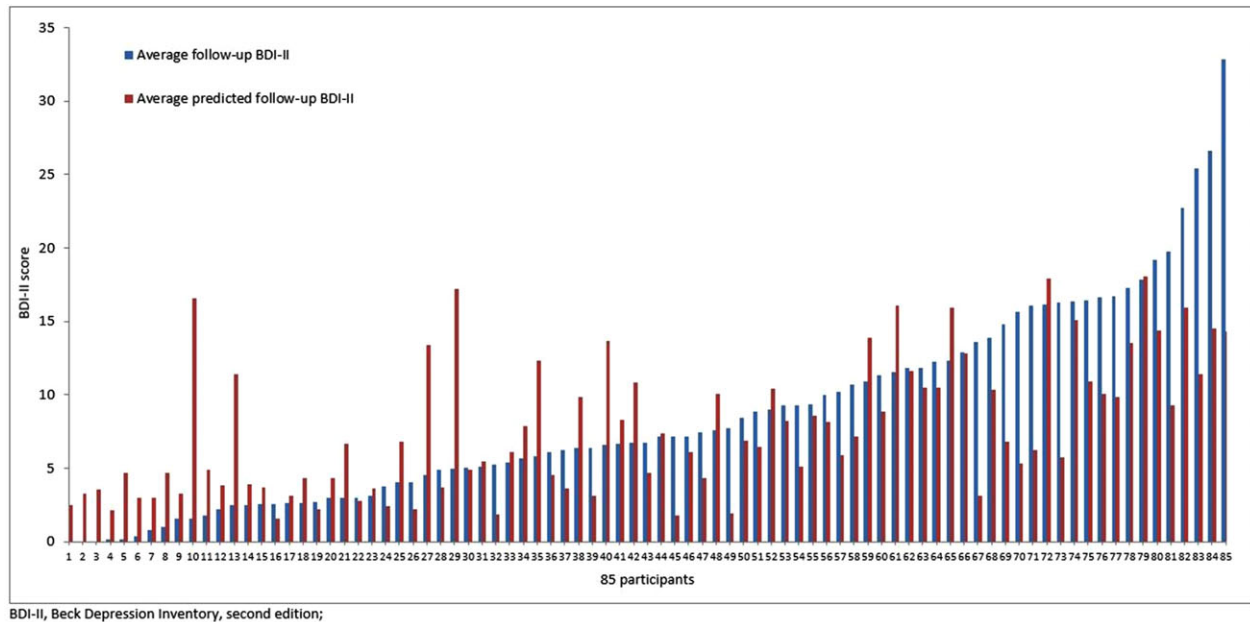
The main goal of this study was to compute a prognostic index (PI) for long-term depression severity after successful psychotherapeutic treatment of MDD. In the context of a recent RCT, we explored 17-month follow-up data of 85 treatment responders to CT and IPT. From the 29 variables examined, three post-treatment predictors (depression severity, hopelessness, and self-esteem) were selected and combined to create a PI. The mean difference between actual and predicted depression severity scores was 0.84, and a total of 21 individuals had a clinically significant difference between these scores. Individual PI scores had a strong correlation with the actual follow-up depression severity, and the association between these severity scores was highly significant.

In this study, we identified three post-treatment predictors for long-term depression severity: post-treatment depression severity, negative beliefs about the future (hopelessness), and lower self-esteem (self-liking and self-competence), all of which have been associated with an unfavorable course. Post-treatment depression severity, or residual depression, has been found to have a strong prognostic value for an unfavorable course of depression (Berwian et al., 2017; Fava, Ruini, & Belaise, 2007; Hardeveld et al., 2010; Nierenberg et al., 2010; ten Doesschate, Bockting, Koeter, & Schene, 2010). Hopelessness as a predictor of long-term depression severity is in line with the hopelessness theory of depression (Abramson, Metalsky, & Alloy, 1989; Liu, Kleiman, Nestor, & Cheek, 2015). This theory hypothesizes that the interaction between negative cognitive styles and negative life events can evoke a sense of hopelessness that in turn contributes to the onset, relapse, and recurrence of depressive episodes. The importance of hopelessness is further supported by the finding that a decrease in hopelessness during various psychotherapy modalities was significantly associated with a decrease in depression severity (Cuijpers et al., 2013). In addition, other studies repeatedly reported hopelessness as a significant predictor of future suicidal behavior and ideation (Beck, Brown, Berchick, Stewart, & Steer, 2006; Kovacs & Garrison, 1985), an important symptom of depression. Self-esteem can be interpreted as an ability that helps individuals to maintain their depressive symptom reduction, conferring resilience. Previous stud-

ies have shown a diminished self-esteem in individuals with depression (Creemers, Scholte, Engels, Prinstein, & Wiers, 2012; Franck, De Raedt, Dereu, & Van den Abbeele, 2007; Phillips, Hine, & Bhullar, 2012) and an increase of self-esteem during various treatment modalities for depression (Barkham et al., 1996; Goldman, Greenberg, & Angus, 2006; Strunk, DeRubeis, Chiu, & Alvarez, 2007). While one study found no predictive effect of post-treatment self-esteem on relapse 1 year after CT (Strunk et al., 2007), others found that an improvement in self-esteem predicted less depressive symptoms after finishing either psychodynamically oriented therapy and cognitive-behavioral group therapy (Allart-van Dam, Hoogduin, & Schaap, 2003; Dinger, Ehrental, Nikendei, & Schauenburg, 2017). Within an interpersonal perspective, low self-esteem and high levels of hopelessness are closely associated with loneliness and low levels of perceived social support which are well known risk factors contributing to the development of MDD (Cacioppo, Grippo, London, Goossens, & Cacioppo, 2015; van Winkel et al., 2017). This cluster of interpersonal factors may potentially serve as targets for further psychotherapeutic or pharmacological treatment in individuals at risk for less-favorable clinical course (Cacioppo et al., 2015).

Based on previous research, we expected the number of previous episodes and childhood maltreatment to be identified as predictors as well (Berwian et al., 2017; Hardeveld et al., 2010; Kovacs et al., 2016; Nanni et al., 2012). However, our selection methods did not identify these variables. Several explanations may be held responsible. First, data on previous episodes and childhood trauma are sensitive to recall-bias, which may result in inconsistent findings between studies. Second, information about previous episodes was gathered by differentiating between "first episode" and "recurrent episode," therefore lumping individuals with two or more episodes together. One could hypothesize that this categorization causes loss of potential predictive information. Third, both previous episodes and childhood maltreatment can be conceptualized as developmental risk factors that may have a mediating role for other predictors to trigger an unfavorable course of depression (Bockting, Spinhoven, Koeter, Wouters, & Schene, 2006; Kendler, Thornton, & Gardner, 2000; Nanni et al., 2012; Segal, Williams, Teasdale, & Gemar, 1996). Following this hypothesis, these variables affect vulnerability for depression through their interaction with more proximal variables such as stressful life events (Kendler et al., 2000), and coping-related factors (Bockting et al., 2006), and were therefore not identified as independent predictors. The findings concerning self-esteem and hopelessness are in line with this, since self-esteem has been hypothesized to come from early attachment experiences (Morley & Moran, 2011) and hopelessness has been hypothesized as a result of interactions between negative cognitive styles and negative life events (Abramson et al., 1989).

After the variable selection phase, the three post-treatment predictors were combined to create a PI, utilizing cross-validated predictions. Individual PI scores predicted long-term depression severity, with a strong correlation with the actual depression severity scores. In addition, for a total of 64 individuals, there was no clinically significant difference between the average actual and predicted depression severity scores, and a permutation test indicated that the association between



**FIGURE 1** Bar plots showing predicted average Beck Depression Inventory (BDI-II) scores and average observed BDI-II scores for each participant during follow-up

the actual and predicted scores was highly significant. Interestingly, responders with higher PI scores had more additional care throughout follow-up, with no significant association between actual depression severity and additional care. The relationship between PI scores and additional support could indicate that the PI scores are also predictive for future care seeking behavior.

We think our variable selection and subsequent PI approach has several strengths. First, instead of focusing on a single predictor, our approach combined multiple predictors and was therefore able to capture the multifactorial nature of the long-term course of MDD (Hardeveld et al., 2013; van Loo et al., 2015; Wang et al., 2014). Second, advanced variable selection methods were applied that can handle multiple predictors and prevent weaker predictors from being overshadowed by stronger ones. Third, the PI scores were calculated by combining the predictors in a meaningful way, with regression coefficients determined by a cross-validation approach. In addition, these regression coefficients were based on a sample that did not include the individual for whom the PI score was estimated for. Our findings differ from earlier studies that investigated multivariate prediction algorithms for the long-term course of MDD (van Loo et al., 2015; Wang et al., 2014). As mentioned earlier, these naturalistic studies do not disentangle treatment effects from a natural course, while this study focuses on post-treatment predictions for long-term outcome after successful controlled treatment.

To our knowledge, this is the first study creating a prognostic index for follow-up outcome after successful and high quality psychotherapy for MDD. Other strengths of this study are the broad range of potential predictors and the multiple outcome measurements throughout follow-up, avoiding recall bias. Our study also has some limitations. First, although we included a broad range of variables covering various domains, there is no guarantee that this is a complete selection of all relevant variables. Second, we did not validate our prediction

model in an independent sample, a so-called “testing dataset.” To minimize overfitting, we did apply a cross-validation approach to our primary model. Moreover, we performed a secondary analysis running the entire process of both variable selection and model-fitting with 5-folds, and this produced an estimated PI-outcome correlation very similar to that obtained in our primary model (0.60 vs. 0.62). Third, throughout follow-up individuals were free to pursue additional treatment. However, the use of additional treatment did not affect the outcome variable and did not seem to have an impact on long-term outcome (Lemmens et al., 2018).

In the long run, PI scores could be used to predict follow-up depression severity for each individual after acute treatment response, to enable clinicians to identify at-risk individuals and personalize long-term treatment strategies. In this context, we propose that the PI would be part of a shared decision-making process, since there is evidence that this would result in higher treatment satisfaction, less dropout, and better clinical outcomes (Lindhiem, Bennett, Trentacosta, & McLear, 2014). Clinicians and patients should collaborate in aggregating and interpreting information on individual treatment goals, residual symptoms, and depression severity predictions (PI scores). As a rule of thumb, a clinically significant difference between post-treatment depression severity and predicted depression severity over time (e.g., a 5 point difference on the BDI-II like in this study), may serve as a meaningful predicted negative prognosis (Hiroe et al., 2005). From a shared decision-making process, informed decisions on maintenance treatment options could be made. Previous studies have shown that psychological interventions (e.g. mindfulness-based), CT and IPT) are suited to prevent relapse and recurrence over 2 years, especially when the individual received acute phase psychotherapy for MDD (Biesheuvel-Leliefeld et al., 2015). However, the PI is not ready for implementation yet, and future research is needed to develop a robust prediction model. First of all, current prediction models are

predominantly composed of risk factors, but accuracy may be enhanced by the inclusion of more resilience factors such as self-esteem (Lorenzo-Luaces, 2015; Waugh & Koster, 2015). Second, prediction models need to be validated in several distinct and related populations. Third, feasibility needs to be considered, as collecting a high number of complex variables could be problematic in routine practice. If replicated, a strength of the current PI is that the three self-report measures can be rapidly assessed in routine practice, most preferably using web-based computerized questionnaires, which enhances feasibility considerably.

## ACKNOWLEDGMENTS

We would like to acknowledge the contribution of participants and therapists at RIAGG Maastricht. Furthermore, we thank Annie Raven and Annie Hendriks for their assistance during the study.

## FUNDING

This research was funded by the research institute of Experimental Psychopathology (EPP), the Netherlands, and the Academic Community Mental Health Centre (RIAGG) in Maastricht, the Netherlands.

## CONFLICT OF INTEREST

The authors declare that they have no competing interests.

## ORCID

Suzanne C. van Bronswijk MD 

<https://orcid.org/0000-0002-2983-1268>

John R. Keefe MA  <https://orcid.org/0000-0001-7250-8960>

## REFERENCES

- Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review*, *96*(2), 358–372.
- Allart-van Dam, E., Hosman, M. C., Hoogduin, C. A., & Schaap, C. P. (2003). The coping with depression course: Short-term outcomes and mediating effects of a randomized controlled trial in the treatment of subclinical depression. *Behavior Therapy*, *34*(3), 381–396.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev. ed.). Washington, D.C.: American Psychiatric Publishing.
- Anderson, K. M., Wilson, P. W., Odell, P. M., & Kannel, W. B. (1991). An updated coronary risk profile. A statement for health professionals. *Circulation*, *83*(1), 356–362.
- Austin, P. C., & Tu, J. V. (2004). Bootstrap methods for developing predictive models. *The American Statistician*, *58*(2), 131–137.
- Barkham, M., Rees, A., Shapiro, D. A., Stiles, W. B., Agnew, R. M., Halstead, J., ... Harrington, V. M. (1996). Outcomes of time-limited psychotherapy in applied settings: Replicating the Second Sheffield Psychotherapy Project. *Journal of Consulting and Clinical Psychology*, *64*(5), 1079–1085.
- Beck, A. T., Brown, G., Berchick, R. J., Stewart, B. L., & Steer, R. A. (2006). Relationship between hopelessness and ultimate suicide: A replication with psychiatric outpatients. *Focus*, *147*(2), 190–296.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.
- Beck, A. T., Steer, R., & Brown, G. K. (1996). *Beck depression inventory II: Manual*. Boston: Harcourt Brace.
- Beck, A. T., & Steer, R. A. (1988). *Manual for the Beck hopelessness scale*. San Antonio: Psychological Corporation.
- Berwian, I. M., Walter, H., Seifritz, E., & Huys, Q. J. (2017). Predicting relapse after antidepressant withdrawal: A systematic review. *Psychological Medicine*, *47*(3), 426–437. <https://doi.org/10.1017/S0033291716002580>
- Biesheuvel-Leliefeld, K. E., Kok, G. D., Bockting, C. L., Cuijpers, P., Hollon, S. D., van Marwijk, H. W., & Smit, F. (2015). Effectiveness of psychological interventions in preventing recurrence of depressive disorder: Meta-analysis and meta-regression. *Journal of Affective Disorders*, *174*, 400–410. <https://doi.org/10.1016/j.jad.2014.12.016>
- Bockting, C. L., Spinhoven, P., Koeter, M. W., Wouters, L. F., & Schene, A. H. (2006). Prediction of recurrence in recurrent depression and the influence of consecutive episodes on vulnerability for depression: A 2-year prospective study. *The Journal of Clinical Psychiatry*, *67*(5), 747–755.
- Burcusa, S. L., & Iacono, W. G. (2007). Risk for recurrence in depression. *Clinical Psychology Review*, *27*(8), 959–985. <https://doi.org/10.1016/j.cpr.2007.02.005>
- Cacioppo, S., Grippo, A. J., London, S., Goossens, L., & Cacioppo, J. T. (2015). Loneliness: Clinical import and interventions. *Perspectives on Psychological Science*, *10*(2), 238–249.
- Chen, J., Pee, D., Ayyagari, R., Graubard, B., Schairer, C., Byrne, C., ... Gail, M. H. (2006). Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute*, *98*(17), 1215–1226. <https://doi.org/10.1093/jnci/djj332>
- Cohen, L., Van den Bout, J., Kramer, W., & Van Vliet, T. (1986). A Dutch attributional style questionnaire: Psychometric properties and findings of some Dutch–American differences. *Cognitive Therapy and Research*, *10*(6), 665–669.
- Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology*, *14*, 209–236.
- Creemers, D. H., Scholte, R. H., Engels, R. C., Prinstein, M. J., & Wiers, R. W. (2012). Implicit and explicit self-esteem as concurrent predictors of suicidal ideation, depressive symptoms, and loneliness. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(1), 638–646.
- Cuijpers, P., de Beurs, D. P., van Spijker, B. A., Berking, M., Andersson, G., & Kerkhof, A. J. (2013). The effects of psychotherapy for adult depression on suicidality and hopelessness: A systematic review and meta-analysis. *Journal of Affective Disorders*, *144*(3), 183–190.
- de Graaf, L. E., Roelofs, J., & Huibers, M. J. (2009). Measuring dysfunctional attitudes in the general population: The dysfunctional attitude scale (form A) revised. *Cognitive Therapy and Research*, *33*(4), 345–355. <https://doi.org/10.1007/s10608-009-9229-y>
- Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy*, *79*, 15–22.
- Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine*, *13*(3), 595–605.
- DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One*, *9*(1), e83875.
- Dinger, U., Ehrental, J. C., Nikendei, C., & Schauenburg, H. (2017). Change in self-esteem predicts depressive symptoms at follow-up after intensive multimodal psychotherapy for major depression. *Clinical Psychology & Psychotherapy*, *24*(5), 1040–1046.



- Fava, G. A., Ruini, C., & Belaise, C. (2007). The concept of recovery in major depression. *Psychological Medicine*, 37(3), 307–317.
- First, M., Spitzer, R., Gibbon, M., & Williams, J. (1995). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)*. New York: Biometrics Research Department, New York State Psychiatric Institute.
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, D.C.: American Psychiatric Press.
- Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Hollon, S. D., Amsterdam, J. D., & Gallop, R. (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *Journal of Consulting and Clinical Psychology*, 77(4), 775–787. <https://doi.org/10.1037/a0015401>
- Franck, E., De Raedt, R., Dereu, M., & Van den Abbeele, D. (2007). Implicit and explicit self-esteem in currently depressed individuals with and without suicidal ideation. *Journal of Behavior Therapy and Experimental Psychiatry*, 38(1), 75–85.
- Garge, N., Eggleston, B., Bobashev, G., & Garge, M. N. (2013). Package 'mobForest'.
- Garge, N. R., Bobashev, G., & Eggleston, B. (2013). Random forest methodology for model-based recursive partitioning: The mobForest package for R. *BMC Bioinformatics*, 14(1), 125.
- Goldman, R. N., Greenberg, L. S., & Angus, L. (2006). The effects of adding emotion-focused interventions to the client-centered relationship conditions in the treatment of depression. *Psychotherapy Research*, 16(5), 537–549.
- Hansen, R., Gaynes, B., Thieda, P., Gartlehner, G., Deveaugh-Geiss, A., Krebs, E., & Lohr, K. (2008). Meta-analysis of major depressive disorder relapse and recurrence with second-generation antidepressants. *Psychiatric Services*, 59(10), 1121–1130. <https://doi.org/10.1176/appi.ps.59.10.1121>
- Hardeveld, F., Spijker, J., De Graaf, R., Hendriks, S. M., Licht, C. M., Nolen, W. A., ... Beekman, A. T. (2013). Recurrence of major depressive disorder across different treatment settings: Results from the NESDA study. *Journal of Affective Disorders*, 147(1–3), 225–231. <https://doi.org/10.1016/j.jad.2012.11.008>
- Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A., & Beekman, A. T. (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica*, 122(3), 184–191. <https://doi.org/10.1111/j.1600-0447.2009.01519.x>
- Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A., & Beekman, A. T. (2013). Recurrence of major depressive disorder and its predictors in the general population: Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Psychological Medicine*, 43(1), 39–48. <https://doi.org/10.1017/S0033291712002395>
- Hiroe, T., Kojima, M., Yamamoto, I., Nojima, S., Kinoshita, Y., Hashimoto, N., ... Furukawa, T. A. (2005). Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression. *Psychiatry Research*, 135(3), 229–235.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56(6), 885–892.
- Hothorn, T., Hornik, K., Van De Wiel, M. A., & Zeileis, A. (2006). A Lego system for conditional inference. *The American Statistician*, 60(3), 257–263.
- Hothorn, T., Hornik, K., Van De Wiel, M. A., & Zeileis, A. (2008). Implementing a class of permutation tests: The coin package.
- Huibers, M. J., Cohen, Z. D., Lemmens, L. H., Arntz, A., Peeters, F. P., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PLoS One*, 10(11), e0140771. <https://doi.org/10.1371/journal.pone.0140771>
- Keefe, J. R., Wiltsey Stirman, S., Cohen, Z. D., DeRubeis, R. J., Smith, B. N., & Resick, P. A. (2018). In rape trauma PTSD, patient characteristics indicate which trauma-focused treatment they are most likely to complete. *Depression and Anxiety*, 35(4), 330–338.
- Kendler, K. S., Thornton, L. M., & Gardner, C. O. (2000). Stressful life events and previous episodes in the etiology of major depression in women: An evaluation of the “kindling” hypothesis. *American Journal of Psychiatry*, 157(8), 1243–1251.
- Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy for depression*. New York: Basis Books.
- Kovacs, M., & Garrison, B. (1985). Hopelessness and eventual suicide: A 10-year prospective study of patients hospitalized with suicidal ideation. *American Journal of Psychiatry*, 1(42), 559–563.
- Kovacs, M., Obrosky, S., & George, C. (2016). The course of major depressive disorder from childhood to young adulthood: Recovery and recurrence in a longitudinal observational study. *Journal of Affective Disorders*, 203, 374–381. <https://doi.org/10.1016/j.jad.2016.05.042>
- Lemmens, L. H., Arntz, A., Peeters, F., Hollon, S. D., Roefs, A., & Huibers, M. J. (2015). Clinical effectiveness of cognitive therapy v. interpersonal psychotherapy for depression: Results of a randomized controlled trial. *Psychological Medicine*, 45(10), 2095–2110. <https://doi.org/10.1017/S0033291715000033>
- Lemmens, L. H., Arntz, A., Peeters, F. P., Hollon, S. D., Roefs, A., & Huibers, M. J. (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy vs. interpersonal therapy for depression: Study protocol for a randomised controlled trial. *Trials*, 12, 150. <https://doi.org/10.1186/1745-6215-12-150>
- Lemmens, L. H., van Bronswijk, S. C., Peeters, F., Arntz, A., Hollon, S. D., & Huibers, M. J. (2018). Long-term outcomes of acute treatment with cognitive therapy v. interpersonal psychotherapy for adult depression: Follow-up of a randomized controlled trial. *Psychological Medicine*, 1–9. [Epub ahead of print]
- Lindhiem, O., Bennett, C. B., Trentacosta, C. J., & McLearn, C. (2014). Client preferences affect treatment satisfaction, completion, and clinical outcome: A meta-analysis. *Clinical Psychology Review*, 34(6), 506–517.
- Liu, R. T., Kleiman, E. M., Nestor, B. A., & Cheek, S. M. (2015). The hopelessness theory of depression: A quarter-century in review. *Clinical Psychology: Science and Practice*, 22(4), 345–365.
- Lorenzo-Luaces, L. (2015). Heterogeneity in the prognosis of major depression: From the common cold to a highly debilitating and recurrent illness. *Epidemiology and Psychiatric Sciences*, 24(6), 466–472. <https://doi.org/10.1017/S2045796015000542>
- Lorenzo-Luaces, L., DeRubeis, R. J., van Straten, A., & Tiemens, B. (2017). A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. *Journal of Affective Disorders*, 213, 78–85.
- Monroe, S. M., & Harkness, K. L. (2011). Recurrence in major depression: A conceptual analysis. *Psychological Review*, 118(4), 655–674. <https://doi.org/10.1037/a0025190>
- Morley, T. E., & Moran, G. (2011). The origins of cognitive vulnerability in early childhood: Mechanisms linking early attachment to later depression. *Clinical Psychology Review*, 31(7), 1071–1082.
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, 180, 461–464.
- Nanni, V., Uher, R., & Danese, A. (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: A meta-analysis. *American Journal of Psychiatry*, 169(2), 141–151. <https://doi.org/10.1176/appi.ajp.2011.11020335>

- Nierenberg, A., Husain, M., Trivedi, M., Fava, M., Warden, D., Wisniewski, S., ... Rush, A. (2010). Residual symptoms after remission of major depressive disorder with citalopram and risk of relapse: A STAR\* D report. *Psychological Medicine*, 40(1), 41–50.
- Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta Earthquake. *Journal of Personality and Social Psychology*, 61(1), 115–121.
- Park, Y., Freedman, A. N., Gail, M. H., Pee, D., Hollenbeck, A., Schatzkin, A., & Pfeiffer, R. M. (2009). Validation of a colorectal cancer risk prediction model among white patients age 50 years and older. *Journal of Clinical Oncology*, 27(5), 694–698. <https://doi.org/10.1200/JCO.2008.17.4813>
- Peterson, C., Semmel, A., Von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. (1982). The attributional style questionnaire. *Cognitive Therapy and Research*, 6(3), 287–299.
- Phillips, W. J., Hine, D. W., & Bhullar, N. (2012). A latent profile analysis of implicit and explicit cognitions associated with depression. *Cognitive Therapy and Research*, 36(5), 458–473.
- Picard, R. R., & Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387), 575–583.
- Raes, F., Hermans, D., & Eelen, P. (2003). Kort instrumenteel De Nederlandstalige versie van de Ruminative Response Scale (RRS-NL) en de Rumination on Sadness Scale (RSS-NL). *Gedragstherapie*, 36, 97–104.
- Rizopoulos, D., & Rizopoulos, M. D. (2009). Package 'bootStepAIC'.
- Segal, Z. V., Williams, J., Teasdale, J., & Gemar, M. (1996). A cognitive science perspective on kindling and episode sensitization in recurrent affective disorder. *Psychological Medicine*, 26(2), 371–380.
- Solomon, D. A., Keller, M. B., Leon, A. C., Mueller, T. I., Lavori, P. W., Shea, M. T., ... Endicott, J. (2000). Multiple recurrences of major depressive disorder. *American Journal of Psychiatry*, 157(2), 229–233. <https://doi.org/10.1176/appi.ajp.157.2.229>
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307.
- Strunk, D. R., DeRubeis, R. J., Chiu, A. W., & Alvarez, J. (2007). Patients' competence in and performance of cognitive therapy skills: Relation to the reduction of relapse risk following treatment for depression. *Journal of Consulting and Clinical Psychology*, 75(4), 523–530.
- Tafarodi, R. W., & Swann, W. B. (2001). Two-dimensional self-esteem: Theory and measurement. *Personality and Individual Differences*, 31(5), 653–673.
- ten Doesschate, C. M., Bockting, C., Koeter, M., & Schene, A. H. (2010). Prediction of recurrence in recurrent depression: A 5.5-year prospective study. *The Journal of Clinical Psychiatry*, 71(8), 984–991.
- Van der Does, W. (2002). Cognitive reactivity to sad mood: Structure and validity of a new measure. *Behaviour Research and Therapy*, 40(1), 105–119.
- van der Zee, K. I., & Sanderman, R. (1993). *Het meten van de algemene gezondheidstoestand met de RAND-36: Een handleiding*. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken.
- van Loo, H. M., Aggen, S. H., Gardner, C. O., & Kendler, K. S. (2015). Multiple risk factors predict recurrence of major depressive disorder in women. *Journal of Affective Disorders*, 180, 52–61. <https://doi.org/10.1016/j.jad.2015.03.045>
- van Winkel, M., Wichers, M., Collip, D., Jacobs, N., Derom, C., Thiery, E., ... Peeters, F. (2017). Unraveling the role of loneliness in depression: The relationship between daily life experience and behavior. *Psychiatry*, 80(2), 104–117.
- Vandromme, H., Hermans, D., Spruyt, A., & Eelen, P. (2007). Dutch translation of the Self-Liking/Self-Competence Scale—Revised: A confirmatory factor analysis of the two-factor structure. *Personality and Individual Differences*, 42(1), 157–167.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4(3), 274–290.
- Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., ... Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, 3(8), e002847.
- Wang, J. L., Patten, S., Sareen, J., Bolton, J., Schmitz, N., & MacQueen, G. (2014). Development and validation of a prediction algorithm for use by health professionals in prediction of recurrence of major depression. *Depress Anxiety*, 31(5), 451–457.
- Waugh, C. E., & Koster, E. H. (2015). A resilience framework for promoting stable remission from depression. *Clinical Psychology Review*, 41, 49–60.
- Weissman, A. N., & Beck, A. T. (1978). Development and validation of the Dysfunctional Attitude Scale: A preliminary investigation.
- Zilcha-Mano, S., Keefe, J. R., Chui, H., Rubin, A., Barrett, M. S., & Barber, J. P. (2016). Reducing dropout in treatment for depression: Translating dropout predictors into individualized treatment recommendations. *The Journal of Clinical Psychiatry*, 77(12), e1584–e1590.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** van Bronswijk SC, Lemmens LHJM, Keefe JR, Huibers MJH, DeRubeis RJ, Peeters FPML. A prognostic index for long-term outcome after successful acute phase cognitive therapy and interpersonal psychotherapy for major depressive disorder. *Depress Anxiety*. 2019;36:252–261. <https://doi.org/10.1002/da.22868>