Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: https://www.tandfonline.com/loi/tpsr20

A placebo prognostic index (PI) as a moderator of outcomes in the treatment of adolescent depression: Could it inform risk-stratification in treatment with cognitive-behavioral therapy, fluoxetine, or their combination?

Lorenzo Lorenzo-Luaces, Natalie Rodriguez-Quintana, Tennisha N. Riley & John R. Weisz

To cite this article: Lorenzo Lorenzo-Luaces, Natalie Rodriguez-Quintana, Tennisha N. Riley & John R. Weisz (2020): A placebo prognostic index (PI) as a moderator of outcomes in the treatment of adolescent depression: Could it inform risk-stratification in treatment with cognitive-behavioral therapy, fluoxetine, or their combination?, Psychotherapy Research, DOI: 10.1080/10503307.2020.1747657

To link to this article: https://doi.org/10.1080/10503307.2020.1747657

Published online: 29 Mar 2020.

Psychotherapy Research, 2020
https://doi.org/10.1080/10503307.2020.1747657

EMPIRICAL PAPER

A placebo prognostic index (PI) as a moderator of outcomes in the treatment of adolescent depression: Could it inform risk-stratification in treatment with cognitive-behavioral therapy, fluoxetine, or their combination?

LORENZO LORENZO-LUACES[1], NATALIE RODRIGUEZ-QUINTANA[1], TENNISHA N. RILEY[12], & JOHN R. WEISZ[3]

[1]*Department of Psychological and Brain Sciences, Indiana University—Bloomington, Bloomington, IN, USA;* [2]*Center for Research on Race and Ethnicity in Society (CRRES), Indiana University—Bloomington, Bloomington, IN, USA &* [3]*Department of Psychology, Harvard University, Cambridge, MA, USA*

(Received 21 August 2019; revised 23 February 2020; accepted 22 March 2020)

Abstract

Introduction: Researchers have proposed that predicting who is a likely placebo responder may help guide treatment allocations to treatment regimens that differ in intensity.

Methods: We used data from the Treatment of Adolescent Depression Study (TADS) in which adolescents ($n = 439$) were randomized 1:1:1:1 to placebo, cognitive-behavioral therapy (CBT), medications (MEDs), or their combination (COMB). We developed a prognostic index (PI) in the placebo group to predict self-reported (RADS) and observer-rated (CDRS) depression outcomes using elastic net regularization. We explored whether the PIs moderated outcomes in the treatment conditions.

Results: PI-CDRS was predicted by multiple variables but it did not moderate outcomes. PI-RADS was predicted by baseline severity, age, sleep problems, expectations, maternal depression, and the action stage of change. It moderated outcomes such that there were treatment differences for less placebo-responsive patients. For participants prone to placebo response, type of treatment had no statistically significant impact on outcomes. Baseline depression severity accounted for this effect: treatment differences were small and non-significant for patients with milder depression but larger in more severely depressed patients.

Discussion: Future work should investigate whether multiple variable explain outcomes beyond severity as well as complex interactions between severity and other variables.

Keywords: personalized medicine; machine learning; stepped care; depression; risk stratification

Clinical or methodological significance of this article: A machine-learning algorithm that identifies placebo responders is useful in guiding treatment assignment to cognitive-behavioral therapy vs. the combination of cognitive-behavioral therapy and medications. Symptom severity was as useful for the purposes of assigning patients to CBT or combination treatment as the multivariable algorithm.

Introduction

Depressive disorders are a leading cause of disability worldwide (Murray et al., 2012). Despite this, there is great variability in the prognosis of individuals who meet the criteria for major depressive disorder (MDD) (Lorenzo-Luaces, 2015; Monroe & Harkness, 2011, 2012). While some individuals have brief depressive episodes that spontaneously recur, others have a chronic or recurrent course. Risk of MDD substantially increases in adolescence, and adolescent-onset MDD is associated with increased chronicity and risk of recurrence (Essau & Chang, 2009; Rohde et al., 2013). Adolescent

Correspondence concerning this article should be addressed to Lorenzo Lorenzo-Luaces, Department of Psychological and Brain Sciences, Indiana University—Bloomington, 1101 E 10th St, Bloomington, IN 47405, USA. Email: lolorenz@indiana.edu

MDD also has a number of adverse developmental consequences, including school drop-out, early pregnancy/parenthood, and long-term unemployment outcomes (Clayborne et al., 2019; Lewinsohn et al., 2003; Rohde et al., 2013).

Among psychological therapies for depression in youth, cognitive-behavioral therapy (CBT) is the most widely researched one (Eckshtain et al., 2019; Weisz et al., 2017; Zhou et al., 2015). Antidepressants are also effective although there are concerns about their risk-benefit ratio (Cipriani et al., 2016). More intensive treatments, including the combination of psychotherapy and antidepressants are available. A challenge for the field has been that of identifying which depressed adolescents will respond well to lower levels of intervention (e.g., CBT only) and which require more intensive treatments (e.g., CBT in combination with medications). In other areas of medicine, like oncology, the problem of how to assign patients to one of two treatment protocols that differ in their intensity is approached by using indices that stratify patients on the basis of their expected prognosis, overall clinical complexity, or risk (Akay et al., 2012; Chen et al., 2004).

Some guidelines for the treatment of depression (NICE, 2019) attempt to risk-stratify patients by making treatment recommendations on the basis of symptom severity. This practice is consistent with finding that symptom severity may be a useful variable in risk stratification across interventions for depression (Bower et al., 2013; Driessen et al., 2010; Fournier et al., 2008). Increasingly, researchers are becoming interested in the use of multiple variables, beyond treatment severity, to optimize treatment allocation (DeRubeis et al., 2014; Lorenzo-Luaces et al., 2017). The idea of risk-stratifying adolescents to different psychosocial interventions by using multiple variables has been explored by Garber and colleagues (Garber et al., 2018; Weersing et al., 2016). These authors identified variables that predicted higher risk of experiencing a depressive episode in adolescents: lower functioning (e.g., significant problems with schoolwork and parents), higher anxiety, and having depressed parents. They divided adolescents who were randomized to receive either a CBT psychoeducational preventative intervention or usual care (i.e., being able to seek mental health services outside the study) into different risk categories. For adolescents at a medium-high level of risk (i.e., lower functioning, high anxiety, or depressed parents), there were no differences between CBT and usual care, suggesting the need for a higher-intensity intervention. For adolescents in the low-risk group, there were large effects of preventive CBT relative to usual care.

Other researchers have used machine learning methods to develop multivariable prognostic indices (PIs) in secondary analyses of adults undergoing treatment for depression and anxiety (Delgadillo et al., 2017; Delgadillo et al., 2016; Lorenzo-Luaces et al., 2017). Machine learning methods are focused on finding patterns and making predictions from data and are well-suited to the task of identifying predictors of outcomes (but, see Makridakis et al., 2018).The multivariable PIs derived from machine learning methods can be considered as continuous indices of the "complexity" of the patient's profile. They array patients on a continuum of overall likelihood of responding to treatment, which can be predicted from existing data. For example, in a sample of patients accessing psychological services, Delgadillo et al. (2017) developed a PI that included demographics (e.g., age), personality factors (e.g., suspiciousness, impulsiveness), and clinical features (e.g., baseline depression, anxiety). Using the PI in another sample, the cases with higher scores on the PI, which the authors interpreted as being more complex cases, had better outcomes if started on a higher-intensity treatment (e.g., a full course of CBT) than a lower-intensity treatment (e.g., guided self-help). No differences were found for patients with a lower PI (i.e., a better expected prognosis). Taken together, the studies using machine learning methods suggest that multiple variables can be combined to predict outcomes in mental health.

Another way to develop a prognostic index, is to model the placebo response (Fournier et al., 2008; Trivedi et al., 2018; Webb et al., 2018). The overall likelihood of responding to placebo relates to patient characteristics that predict naturalistic course of illness (Rutherford & Roose, 2013). The high placebo response in MDD may make it difficult to ascertain relative differences between treatment approaches, even those that differ in their intensity (e.g., CBT only vs. combination treatment). It has been stated that identifying likely placebo responders could lead to optimization of outcomes by allocating patients who are likely to respond to placebos to less-intensive interventions and reserving high-intensity treatment for patients who are not predicted to respond to placebos (Fournier et al., 2008; Trivedi et al., 2018; Webb et al., 2018). Recently, machine learning methods have revealed various predictors of depressed patients' response to placebo including depression severity, age, history of abuse, and other variables (Trivedi et al., 2018; Webb et al., 2018). Webb et al. (2020) recently compared various machine-learning approaches to predict outcomes in a naturalistic sample and found that elastic net

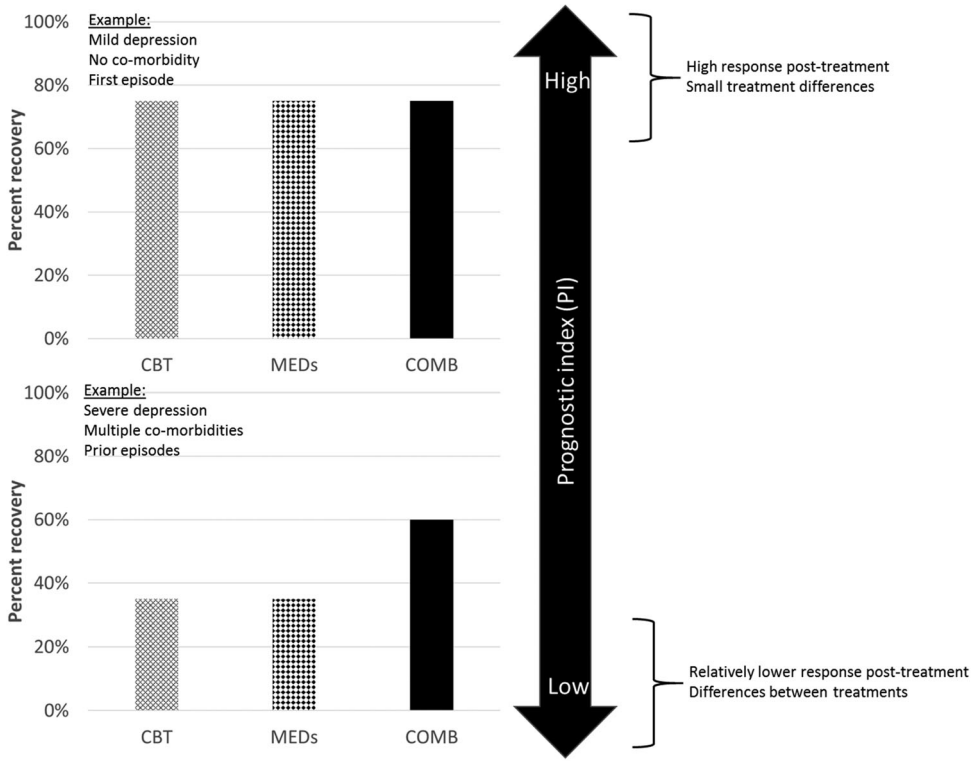

Figure 1. Hypothesized treatment differences between cognitive-behavioral therapy (CBT), the antidepressant medication fluoxetine (ADM), or their combination across levels of a prognostic index (PI) predicting placebo responsiveness.

regularization outperformed other methods. Elastic net is an attractive approach to predict outcomes because it performs variable selection and regularization. Moreover, unlike other "black box" approaches to machine-learning, elastic net produces a linear model that will be familiar to user of linear regression.

Although there is evidence to support risk-stratification in mental health care, as well as studies that have explored predictors of placebo response, to our knowledge no previous study has developed a placebo PI and explored its utility in risk-stratification. The objective of the current study is to use data from the Treatment for Adolescent Depression Study (TADS) (March et al., 2004) to develop a prognostic index (PI) that can be used to assign adolescent patients to CBT or medication monotherapy vs. combination treatment. The TADS dataset is ideally-suited to the task of exploring the utility of a placebo PI because it is one of the few studies that randomized patients to a placebo condition, two therapies delivered singly, and their combination. We utilized elastic net regularization for our prediction efforts. We hypothesized that differences between COMB vs. CBT and antidepressant medications (MEDs) would be most pronounced among those who are least likely to be placebo-responsive (see Figure 1).

Methods

Treatment for Adolescents with Depression Study (TADS)

TADS was a multicenter, randomized controlled trial that evaluated the efficacy of an acute (i.e., 12-week) phase of: CBT, MEDs (i.e., fluoxetine), their combination, and a pill placebo for adolescents with MDD. Several articles have discussed its rationale, design, and methods (March et al., 2004; TADS Team, 2003, 2005). Participants consisted of 439 adolescents, aged 12–17 that had primary MDD. MDD diagnosis was determined by an independent evaluator using the Schedule for Affective Disorders and Schizophrenia for School-Aged Children-Present and Lifetime Version (Kaufman et al., 1997). In primary analyses, which used last-observation carried forward (LOCF) imputation, response rates according to the Clinical Global Impression (CGI) were: 71.0% in combination treatment, 60.6% in fluoxetine, 43.2% in CBT, and 34.8% with placebos.

Treatment Conditions

Cognitive behavior therapy (CBT, $n = 111$). CBT was delivered as a skills-based treatment for which the core assumption is that depression is

Table 1. Demographic characteristics of adolescents in the Treatment for Adolescent Depression Study (TADS; $n = 439$) with missing or imputed data.

	Unimputed			Imputed		
	M	*SD*	% miss	*M*	*SD*	SMD
CDRS						
Baseline	59.98	10.09	20.0%	59.97	10.10	0.00
Mid-treatment	42.21	11.88	12.0%	42.11	12.27	0.01
End of treatment	38.56	13.01	14.0%	38.27	13.45	0.02
RADS						
Baseline	79.18	13.94	11.0%	79.28	14.44	−0.01
Mid-treatment	66.31	15.28	26.0%	66.82	16.72	−0.03
End of treatment	62.32	15.54	20.0%	62.10	16.54	0.01
Age	14.61	1.54	0.0%	14.61	1.54	0.00
Functioning (CGAS)	49.65	7.29	3.0%	49.64	7.37	0.00
Suicidality (SIQ)	23.81	21.46	6.0%	23.68	21.95	0.01
Depression duration (weeks)	69.37	85.04	0.0%	69.37	85.04	0.00
Treatment expectations (TEA)	2.37	0.64	9.0%	2.37	0.67	0.01
Action stage (SOC)	13.64	2.38	7.0%	13.65	2.47	0.00
Maternal BDI	12.71	9.29	11.0%	12.63	9.83	0.01
Somatic anxiety (MASC)	6.25	3.86	3.0%	6.25	3.90	0.00
Life stress (PLES)	3.73	3.46	10.0%	3.69	3.63	0.01
Vocabulary IQ (WISC)	10.61	2.45	4.0%	10.61	2.50	0.00
Sleep problems (CDRS)	3.71	1.52	0.0%	3.71	1.52	0.00
Sleep problems (RADS)	2.95	1.05	3.9%	2.95	1.07	0.00
Illness or disability (HON 6)	2.10	1.00	1.0%	2.10	1.00	0.00
Social relations (HON 10)	0.34	0.47	1.0%	0.34	0.47	0.00
	n	%	% miss	*n*	%	%diff
Female (vs. Male)	239	54.44%	0.0%	239	54.44%	0.00%
White (vs. non-White)	355	80.87%	0.0%	355	80.87%	0.00%
Parents live together (yes/no)	250	56.95%	0.2%	250	56.95%	0.00%
Serious trauma (yes/no)	122	27.79%	0.0%	122	27.79%	0.00%
Anxiety co-morbidity (yes/no)	113	25.74%	10.9%	113	25.74%	0.00%
Income			3.2%			
0–$19,000	236	53.76%		244	55.58%	−1.82%
$20,000–$39,999	101	23.01%		103	23.46%	−0.46%
$40,000 and over	88	20.05%		92	20.96%	−0.91%

Note: *M* = mean, *SD* = standard deviation, % miss = percentage of missing data, SMD = standardized mean difference between imputed and unimputed data, % diff = absolute difference in percentage between imputed and unimputed data, Reynolds Adolescent Depression Scale (RADS), Children's Depression Rating Scale (CDRS), Children`s Global Assessment Scale (CGAS), Suicidal Ideation Questionnaire (SIQ), Average Treatment Expectancy across CBT, ADM, and COMBO, Stages of Change (SOC), Beck Depression Inventory-II (BDI), Mulitdimensional Anxiety Scale for Children (MASC), Pediatric Life Events Screen (PLES), Health of Nations Scale (HON).

maintained by lack of positive reinforcement from the environment and negative thought patterns. Adolescents met with a therapist weekly for 50–60 minutes.

Medication (MEDs, $n = 109$). Adolescents assigned to the medication condition received Fluoxetine; the initial dosage consisted of 10 mg/day and increased of necessary up to 40 mg/day in Stage 1 and 60 mg/day in Stage 2. Youth met with a pharmacotherapist, who provided medication management and encouragement, for 20–30 minutes.

Combination (COMB, $n = 107$). Adolescents assigned to the combination group received all the components of both CBT and medication (as described above).

Placebo (PBO, $n = 112$). Adolescents assigned to the placebo group received a pill placebo instead of fluoxetine. Dosage and meeting structure were equivalent to the medication group.

Outcome Measures

TADS used CGI ratings as their outcome and also included continuous measurement of depression with the Reynolds Adolescent Depression Scale (RADS; Reynolds, 1987, 2004) and Children's Depression Rating Scale-Revised (CDRS-R; Poznanski & Mokros, 1996). We used these two measures as our outcomes, given that prior research has cast doubts on the reliability and validity of CGI ratings (Ruhé et al., 2005).

RADS. The RADS (Reynolds, 1987, 2004) is a 30-item self-reported depression severity scale that was completed at baseline, mid-treatment (i.e., week 6), and end of treatment (i.e., week 12). Scores range from 30–120 with higher scores indicating more severe symptoms. Research supports its validity and reliability in measuring adolescent depression (Reynolds, 2004). In the current sample, it appeared to be a reliable measure of depression (alpha = 0.91).

CDRS-R. The CDRS-R (Poznanski & Mokros, 1996) is a 17-item depression severity scale completed by blinded independent evaluators at baseline, mid-treatment (i.e., week 6), and end of treatment (i.e., week 12). Scores range from 17 to 113 with higher scores indicating more severe depressive symptoms. Research supports its validity and reliability in measuring adolescent depression (Mayes et al., 2010). In an evaluation of CDRS-R ratings among 20% of the sample, the intraclass correlation coefficient (ICC) for the CDRS-R was 0.95, suggesting excellent interrater reliability.

Missing Data

At baseline, rates of missing data on baseline co-variates we selected were low (all <12%, see Table 1). Rates of missing outcome data were somewhat higher (e.g., 14% CDRS at end of treatment and 26% for missing RADS scores mid-treatment). Missingness in mid-treatment RADS appeared to be somewhat lower in the medication condition (18%) than CBT (32%), COMB (28%), or placebo (24%) but this effect was not statistically significant ($x^2(3) = 5.54$, $p = 0.14$). There were no other indications that missingness differed between the treatments (ps > 0.38). As would be expected given randomization, rates of missingness on the baseline co-variates did not vary between treatments (ps > 0.24).

To address missing data, including missing outcome assessments, we used a machine learning algorithm: non-parametric missing value imputation using random forests, with the R package missForest (Stekhoven & Bühlmann, 2012). Before imputation, we winsorized variables that had outliers at three standard deviations over or under the mean. Two sets of imputation trees were run separately: one for the placebo condition (which will be trained on for prediction purposes) and one for the treatment conditions (in which the utility of the predictions will be tested). Multivariable imputation models are preferred to complete case analyses, which are less representative of the sample, have less power, and are subject to bias due to the nature of missingness (Janssen et al., 2009; Moons et al., 2006; Sterne et al., 2009; van Kuijk et al., 2016). As well, they are preferable to analyses with LOCF imputations (Kenward & Molenberghs, 2009; Lachin, 2016).

The variables in the imputation model included the baseline co-variates reported in Table 1, outcome data, as well as related auxiliary variables and outcomes over time (see Supplement 1). It is sometimes assumed that imputation models should not include the study outcome, opting for LOCF imputation. However, this assumes that the relationship between the outcomes and other variables is 0 and the empirical literature on this topic suggests that omitting outcomes from imputation analyses actually can lead to biased statistical estimates at worst (Moons et al., 2006; Young & Johnson, 2010) or is inert at worst (Johnson & Young, 2011). Imputations based on machine learning, like the random forest, have been recommended based on their performance in predictions efforts (Jerez et al., 2010; Stekhoven & Bühlmann, 2012; Waljee et al., 2013).

Analytic Plan

Analyses were conducted using the R programing language (R Core Team, 2020). Our outcomes were the week 12 RADS and CDRS scores, which would have been the end-of-treatment outcomes for patients who remained in the trial and the imputed outcomes for those who dropped out. To determine which variables to use for prediction purposes, we searched the TADS dataset for variables identified by Kessler et al. (2016) to predict outcomes, including demographics, depression features, co-morbid symptoms, stress history, and other features. We also searched adolescent-specific reviews of predictors of outcomes which highlighted the role of maternal depression, treatment expectancies, and readiness to change (Curry et al., 2006). Overall, we identified 18 variables that were plausible candidate predictors of outcome useful for risk stratification. All variables were standardized for analysis.

To determine which of the 18 variables would be included in our PI, we ran an elastic net regularization (ENR) model in the placebo group with 10-fold cross-validation. ENR belongs to the family of penalized regression procedures, combining the L1 and L2 penalizations that are characteristic of the LASSO and ridge regression (Hastie et al., 2009). The method shrinks coefficients in the final model so that the estimates are more likely to replicate than estimates from other procedures like ordinary least squares regression. Variables that have very small coefficients have them set to zero so the procedure also functions for variable selection. Penalty

terms for alpha and lambda were selected by running 1000 bootstraps of the analyses, and determining which penalty terms were associated with the lowest mean square error (MSE).

Once a PI was developed, by determining which variables predict outcome in the placebo group, we focused our analyses on the CBT, MEDs, and COMB conditions. For each patient in each of the three treatment arms, we compute a PI. This is considered an index of how "placebo-responsive" each patient would be expected to be, had they been assigned to the placebo condition. To determine whether the PIs moderated outcome differences between the monotherapies (i.e., MEDS and CBT) relative to the treatment conditions, we ran regression models in which the depressive symptoms were the outcomes (i.e., RADS and CDRS) and the predictor variables were dummies for CBT (vs. COMB) or MEDS (vs. COMB), the PIs, and the interaction of the PIs with the treatment contrasts. Thus, for example, the model exploring whether the PI moderated outcomes for the RADS was as follows:

$$\begin{aligned}\mathrm{RADS}_{12} = {} & \beta 0 + (\beta 1 * \mathrm{CBT}) + (\beta 2 * \mathrm{MEDS}) \\ & + (\beta 3 * \mathrm{PI}_{\mathrm{RADS}}) + (\beta 4 * \mathrm{MEDS} * \mathrm{PI}_{\mathrm{RADS}}) \\ & + (\beta 5 * \mathrm{CBT} * \mathrm{PI}_{\mathrm{RADS}})\end{aligned}$$

In this model, the terms for $\beta 4$ and $\beta 5$ indicate whether the coefficient for the interaction between the $\mathrm{PI}_{\mathrm{RADS}}$ and CBT and the $\mathrm{PI}_{\mathrm{RADS}}$ and MEDS is significantly different from zero (see Table 3). If the tests associated with these coefficients are statistically significant, that indicates that outcomes vary between the monotherapy and combined treatment across levels of the PI. We conducted a similar regression comparing CBT vs. MEDS. To probe significant interactions between the PIs and treatment conditions, we used the Johnson-Neyman (J-N) technique (Hayes & Matthes, 2009). The J-N technique can be used to identify the point at which the relationship between two variables (e.g., treatment condition and depression outcomes) is altered by a third variables (e.g., the PI). In our analyses, the J-N critical significance region would indicate at what level of placebo responsiveness treatment differences emerge or disappear.

Results

As would be expected given the results obtained in original publication with the LOCF imputed dataset, the best outcomes were obtained in the COMB condition (RADS: $M = 56.73$, $SD = 15.59$, CDRS: $M = 34.09$, $SD = 11.81$), followed by MEDs (RADS: $M = 60.51$, $SD = 15.84$, CDRS: $M = 36.73$, $SD = 12.33$), CBT (RADS: $M = 66.35$, $SD = 15.57$, CDRS: $M = 41.06$, $SD = 13.62$), and placebo (RADS: $M = 65.43$, $SD = 13.33$, CDRS: $M = 42.13$, $SD = 12.71$). As in the LOCF-imputed data, the COMB ($ps < 0.001$) and MEDs ($ps < 0.02$) but not CBT ($ps > 0.53$) conditions proved superior to placebo.

Reynolds Adolescent Depression Scale (RADS)

Of the 18 predictor variables submitted to the Elastic Net procedure, 6 were retained as predictors of response in the placebo condition. Baseline depression severity, age, higher endorsement of the action stage of change, sleep problems, and worse expectations of improvement predicted higher post-treatment depression scores, while maternal depression severity predicted lower post-treatment depression scores. Table 2 shows the coefficients assigned to these variables by the Elastic Net model. For the interested reader, we also present the results obtained when the variables are entered into a single ordinary least squares (OLS) regression predicting end of treatment RADS. As can be seen in Table 2, the ENR model shrinks the regression coefficients so they are more conservative; the coefficients from the ENR that are used to generate the PI (mean squared error (MSE) = 152.83).

Within the placebo group, the resulting $\mathrm{PI}_{\mathrm{RADS}}$ ($M = 65.43$, $SD = 5.69$), generated from the ENR, evidenced fair predictive accuracy ($R^{2\mathrm{pred}} = 0.33$, 95% CI = 0.20–0.48, $p < 0.001$, normalized root mean square error (NRMSE) = 0.19). We generated $\mathrm{PI}_{\mathrm{RADS}}$ scores for every patient in the TADS dataset (i.e., outside the placebo group). For the full sample, the PI ($M = 64.41$, $SD = 5.87$) was a fair predictor of depression outcomes ($R^{2\mathrm{pred}} = 0.19$, 95% CI = 0.11–0.26, $p < 0.001$, NRMSE = 0.18). A greater score on the $\mathrm{PI}_{\mathrm{RADS}}$ indicates patients were predicted to have relatively poorer outcomes (i.e., to be less placebo-responsive). A lower score indicates patients were predicted to have relatively better outcomes (i.e., to be more placebo-responsive). There were no differences between the three actives treatments in the levels of the $\mathrm{PI}_{\mathrm{RADS}}$ ($F(2, 324) = 1.04$, $p = 0.36$).

There was an interaction between the CBT (vs. combination treatment) condition and the $\mathrm{PI}_{\mathrm{RADS}}$ in predicting outcomes, but no interaction between the MEDS (vs. combination treatment) condition and the $\mathrm{PI}_{\mathrm{RADS}}$ in predicting outcomes (see Table 3). As hypothesized, the nature of interaction was such that differences between the combination

Table 2. Prediction of end of treatment depression in adolescents randomzied to placebo ($n = 112$) with elastic net regularization (ENR) or ordinary least squares (OLS) regression.

RADS	***B* (ENR)**	***B* (OLS)**	***SE***	***t***	***β***
Intercept	64.67	64.50	1.04	62.18	***
Action stage (SOC)	0.65	2.21	1.04	2.12	0.17*
Maternal BDI	−0.56	−1.93	0.98	−1.96	−0.15
Age	1.22	2.04	1.00	2.05	0.16*
Treatment expectations (TEA)	0.02	1.33	1.04	1.28	0.10
Sleep problem (CDRS/RADS)	0.18	1.29	1.06	1.22	0.10
Baseline severity (RADS)	5.54	6.97	1.08	6.44	0.51***
CDRS	***B* (ENR)**	***B* (OLS)**	***SE***	***t***	***β***
Intercept	42.15	42.22	1.17	36.16	***
Duration depression (weeks)	0.36	1.15	1.24	0.93	0.09
Treatment expectations (TEA)	0.54	0.57	1.23	0.46	0.05
Action stage (SOC)	−0.06	−0.18	1.22	−0.15	−0.01
Maternal BDI	−0.68	−2.46	1.18	−2.08	−0.20*
Somatic anxiety (MASC)	0.25	1.64	1.39	1.19	0.12
Age	1.06	3.09	1.21	2.56	0.26*
Pediatric Life Events Screen (PLES)	−0.11	−0.10	1.24	−0.08	−0.01
Parents live together	−0.05	0.71	1.27	0.56	0.06
Income	0.30	0.71	1.38	0.51	0.05
IQ (WAIS—Vocabulary)	−0.17	−0.80	1.23	−0.65	−0.06
Sleep problem (CDRS/RADS)	0.04	0.65	1.30	0.50	0.05
Suicidality (SIQ)	−0.63	−3.10	1.27	−2.44	−0.24*
Serious trauma	−0.21	0.34	1.24	0.27	0.03
Any anxiety diagnosis	−0.35	−1.30	1.31	−1.00	−0.10
Baseline severity (CDRS)	1.06	4.70	1.51	3.10	0.37**
Illness or disability (HON 6)	−0.23	−0.73	1.20	−0.61	−0.06
Social relations (HON 10)	−0.02	−0.36	1.30	−0.28	−0.03
Functioning (CGAS)	0.27	3.23	1.52	2.13	0.26*

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, | | $p < 0.10$; Reynolds Adolescent Depression Scale (RADS), Children's Depression Rating Scale (CDRS), Children`s Global Assessment Scale (CGAS), Suicidal Ideation Questionnaire (SIQ), Average Treatment Expectancy across CBT, ADM, and COMBO, Stages of Change (SOC), Beck Depression Inventory-II (BDI), Mulitdimensional Anxiety Scale for Children (MASC), Pediatric Life Events Screen (PLES), Health of Nations Scale (HON).

treatment and the CBT monotherapy were most pronounced among patients who were predicted to be less responsive to placebo. The Johnson-Neyman technique (see Figure 2, panels A and B) suggested that at a $PI_{RADS} \geq 58.64$ (i.e., less placebo-responsive) CBT was less effective than COMB ($B = 11.06$, $SE = 2.12$, $t = 5.21$, $p < 0.0001$, SMD = 0.68, 95% CI = 0.44–0.91). Within this value of the PI_{RADS}, MEDS were also superior to CBT ($B = -6.27$, $SE = 2.09$, $t = -3.01$, $p = 0.003$, SMD = 0.44, 95% CI = 0.20–0.67). By contrast, a PI_{RADS} score below 58.64 (i.e., model predicted better placebo response) indicated no statistically significant difference between CBT and COMB ($B = 4.82$, $SE = 3.86$, $t = 1.25$, $p = 0.22$, SMD = 0.35, 95% CI = −0.36–1.06). Below this cut-off, MEDs were not superior to CBT ($B = 2.05$, $SE = 3.48$, $t = 0.59$, $p = 0.56$, SMD = −0.04, 95% CI = −0.52–0.43).

Children's Depression Rating Scale (CDRS)

Of the 18 predictor variables submitted to the Elastic Net procedure, all were retained as predictors of response in the placebo condition, though most were weak predictors. The strongest loadings were from baseline severity, age, treatment expectancies, and having parents who lived together. Table 2 shows the weights assigned to these variables by the Elastic Net model (MSE = 163.32). Within the placebo group, the resulting PI_{CDRS} ($M = 42.13$, $SD = 2.09$), generated from the ENR, evidenced fair predictive accuracy ($R^{2pred} = 22\%$, 95% CI = 0.09–0.36, $p < 0.001$, NRMSE = 0.21). Out of sample, the PI_{CDRS} ($M = 42.16$, $SD = 1.88$) was a fair predictor of depression outcomes ($R^{2pred} = 2\%$, 95% CI = 0.001–6%, $p = 0.009$, NRMSE = 0.22). There were no differences between the three actives treatments in the levels of the PI_{CDRS} ($F(324,2) = 0.30$, $p = 0.74$). Contrary to our hypothesis, there was no interaction between the CBT (vs. combination treatment) condition and the PI_{CDRS} in predicting outcomes nor was there an interaction between the MEDS (vs. combination treatment) condition and the PI_{CDRS} in predicting outcomes (see Table 3).

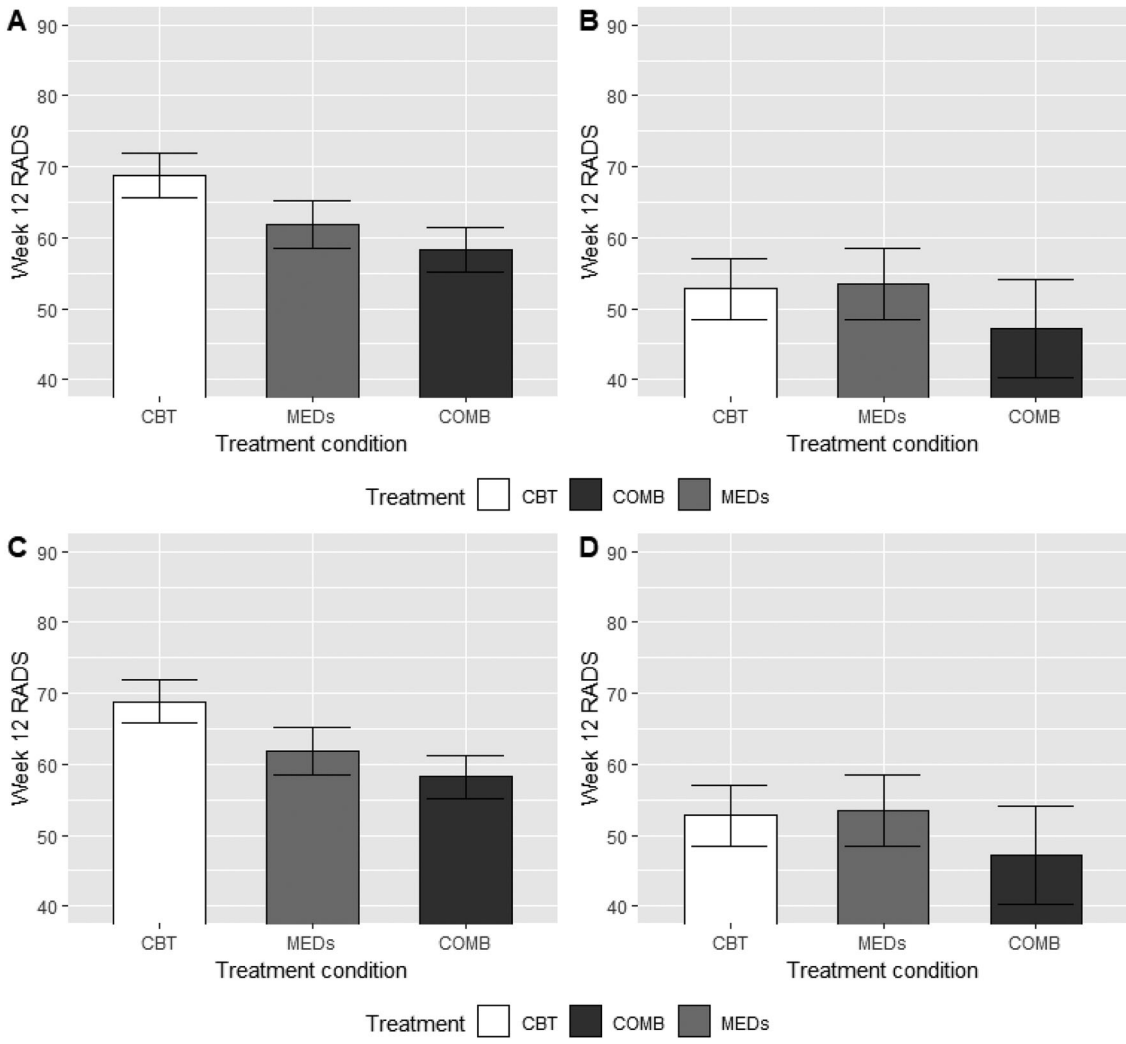

Figure 2. Treatment differences between depressed adolescents ($n = 327$) randomized to cognitive-behavioral therapy (CBT), the antidepressant medication fluoxetine (MEDs), or their combination (COMB) by level of prognostic index (PI) for the Reynolds Adolescent Depression Scale (RADS).
Note. A = Patients with a multivariable prognostic index (PIRADS) ≥ 58.64, B = Patients with a PIRADS < 58.64, C = Patients with a single variable PI based on symptom severity alone, with a predicted score ≥56.48, D = Patients with a single variable PI based on symptom severity alone, with a predicted score <56.48.

Sensitivity Analyses

The PI-RADS proved to be a moderator of outcomes. Inspections of the variables that compose this algorithm, as well as their effect sizes revealed that baseline severity was the strongest contributor to the model (see Table 2). To assess whether the utility of the PI-RADS could simply be accounted for by the effects of baseline severity, we reran all analyses without allowing symptom severity to be included in the variable selection process. The resulting PI (MSE = 176.19) was still related to outcomes within the placebo sample ($R^2 = 30\%$, 95% CI = 0.17–0.44, $p < 0.001$, NRMSE = 0.20) and predicted outcomes in the treatment samples ($R^{2\text{pred}} = 9\%$, 95% CI = 0.04–0.15, $p < 0.001$, NRMSE = 0.19). However, this PI-RADS, devoid of severity, did not moderate outcomes in CBT (vs. combination treatment) ($B = 0.20$, $SE = 0.44$, $t = 0.47$, $p = 0.64$) or MEDS (vs. combination treatment) ($B = 0.07$, $SE = 0.47$, $t = 0.14$, $p = 0.89$). By way of contrast, a PI constructed in the placebo sample using only baseline moderated outcomes in CBT (vs. combination treatment) ($B = 0.55$, $SE = 0.27$, $t = 2.05$, $p = 0.04$) but not MEDs (vs. combination treatment) ($B = -0.14$, $SE = 0.27$, $t = -0.54$, $p = 0.59$).

Using the Johnson-Neyman technique revealed that for patients who were expected to be placebo-responsive, by virtue of having a less severe depression that predicted a RADS <56.48, the

Table 3. Effects of prognostic indices (PIs) on treatment outcomes in cognitive-behavioral therapy (CBT) or medications (MEDs) vs. their combination ($n = 337$).

RADS	***B***	***SE***	***t***
Intercept	63.59	0.94	67.85
CBT (vs. COMB)	9.90	1.88	5.26***
MEDs (vs. COMB)	4.82	1.90	2.54*
PI	1.31	0.16	8.35***
CBT * PI_{RADS}	0.78	0.33	2.35*
ADM * PI_{RADS}	−0.13	0.32	−0.39
CDRS	**B**	***SE***	***t***
Intercept	38.88	0.84	46.25
CBT (vs. COMB)	6.96	1.69	4.12***
MEDs (vs. COMB)	2.72	1.70	1.60
PI	1.10	0.44	2.51*
CBT * PI_{CDRS}	1.41	0.90	1.56
ADM * PI_{CDRS}	−0.35	0.93	−0.38

Note: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, COMB = Combination of CBT and medication, PI = prognostic index, RADS = Reynolds Adolescent Depression Scale, CDRS = Children's Depression Rating Scale.

differences in outcomes between COMB and CBT ($B = 7.84$, $SE = 4.58$, $t = 1.71$, $p = 0.09$, SMD = 0.50, 95% CI = −0.26–1.25) were not statistically significant nor were the differences between CBT and MEDs ($B = 1.48$, $SE = 3.74$, $t = 0.40$, $p = 0.70$, SMD = −0.09, 95% CI = −0.75–0.56). For patients who were expected to be less placebo-responsive, by virtue of having a predicted severity score RADS ≥ 56.48, COMB was superior to CBT ($B = 10.61$, $SE = 2.25$, $t = 4.71$, $p < 0.001$, SMD = 0.68, 95% CI = 0.38–0.98) as were MEDS ($B = -6.97$, $SE = 2.31$, $t = -3.01$, $p = 0.003$, SMD = 0.44, 95% CI = 0.15–0.75). Taken together, these results suggest that the utility of the multivariable PI-RADS in predicting outcomes was largely accounted for by baseline depression severity. Intake severity on the RADS, alone, explained 18% of the variance in RADS outcomes (95% CI 0.11–0.25, $p < 0.001$). Adding the interaction of severity and the treatment conditionss increased this ($\Delta R^2 = 0.08$) and this change was statistically significant ($p < 0.001$). Similarly, our PI_{RADS} and its interaction with treatment explained variance above and beyond severity by itself ($\Delta R^2 = 0.08$, $p < 0.001$) but not above and beyond severity and its interactions with treatments ($\Delta R^2 = 0.01$, $p = 0.27$).

Discussion

We developed a PI to predict the placebo response in adolescents with depression. We then explored whether this PI was a moderator of outcomes in combination treatment versus monotherapy with fluoxetine or CBT. In prior investigations of this kind, researchers have assumed that being able to identify placebo responders should assist in guiding treatment decisions (Trivedi et al., 2018). For the RADS, we found that lower levels of expected placebo response predicted larger treatment differences between COMB vs. CBT monotherapy, and superior outcomes for MEDs compared to CBT. For patients with relatively low PIs (i.e., those who were expected to be placebo responders), no differences between COMB, MEDs, and CBT were found. This was consistent with our hypothesis. Notably, baseline severity largely accounted for this effect. Contrary to our prediction, the PI developed on the CDRS was not a moderator of outcomes. Taken together, these results may suggest that the application of multivariate PIs for placebo response may work better for some measures than others (in the present case, self-report), and may in some cases be accounted for by single variables (in the present case, baseline severity).

Interestingly, our models were better able to predict self-reported depression with the RADS than observer-reported depression with the CDRS and only the former PI moderated outcomes. There are several possible reasons for this. First, baseline severity on the RADS ($r = 0.43$, 95% CI = 0.35–0.50) was more predictive of end-of-treatment RADS than baseline CDRS was predictive of outcomes ($r = 0.24$, 95% CI = 0.14–0.35), and this variable proved to be carrying most of the variance in risk stratification. Additionally, differences in measures of depression have been noted in relation to item content (Fried, 2017). Here, while the items overlap substantially, the RADS oversamples items related to positive affectivity and also includes items tapping worries, negative judgements, and social withdrawal that are not in the CDRS, which instead measures "objective" indicators of depression (e.g., tempo of speech) not in the RADS.

Limitations

An important limitation of our analysis is that the current sample is composed entirely of adolescents, thus limiting the generalizability of our results. It is possible, for example, that symptom severity is the strongest variable predicting outcomes in adolescents whereas, in adults, variables that go beyond symptom severity (e.g., unemployment status, sleep) are stronger predictors of outcomes (Lorenzo-Luaces et al., 2017). It is also possible that our prediction efforts were limited by the variables available in the TADS dataset. Missing from that dataset, for example, were variables relevant to temperament and to social context, both of which could influence prediction of treatment outcomes in a sample of

adolescents. Reviews of the literature also suggest that studies that combine self-report, neuroimaging, and genetic methods yield better prediction accuracy than studies that rely on self-report alone (Lee et al., 2018). The combination of these types of variables, which were not measured in TADS, could have led to more precise risk stratification in the current study. An addition limitation was sample size; although the TADS sample was unusually large for a depression treatment study (c.f.,Barth et al., 2016), it was nonetheless underpowered for differential prediction efforts (Luedtke et al., 2018). Specifically, Luedtke et al. (2018) have shown that power in multivariate prediction problems is a function of the improvement that using a variable like a PI would have on overall outcomes, the noise in the measurement of the variables, and the overall remission rate. In a sample like ours, the use of the PI would have to result in a modest increase in outcomes (>15%) to have acceptable power (i.e., around 80%) to detect a statistically significant interaction at $p < 0.05$. Future research with larger samples should test the utility of PIs, especially those that may have small effects.

Despite these limitations, to our knowledge, this is the first study to directly test the expected utility of using the expected placebo response in assigning patients to one of two monotherapies vs. combination treatment. We employed techniques from machine-learning, which are devised to minimize the risks of overfitting to create our prediction model. The study thus serves as a potentially useful proof-of-concept investigation, illustrating a general strategy and machine-learning application that researchers may add to their toolbox in efforts to guide treatment assignment.

Implications

There is widespread recognition that the field needs to move beyond a "one-size-fits-all" approach to the treatment and prevention of depression and multiple other mental health conditions (Cohen & DeRubeis, 2018; Kessler, 2018; Kessler et al., 2016). Multivariable risk-stratification based on prognosis to different psychosocial interventions has previously received support in the literature on prevention of depression with CBT psychoeducation in adolescents (Garber et al., 2018; Weersing et al., 2016), treatment of adult depression (Lorenzo-Luaces et al., 2017), and treatment of mixed anxiety-depression samples (Delgadillo et al., 2016, 2017). Our results are part of the broad enterprise of multivariable risk-stratification and call attention to the importance of making strong tests of the utility of PIs, going beyond whether they simply predict outcomes above and beyond chance. Some (Lorenzo-Luaces et al., 2017; Webb et al., 2018) but not all (Delgadillo et al., 2016, 2017) of the studies providing support for the utility of a PI have tested a multivariable algorithm against simpler solutions like symptom severity.

Indeed, we found support for the idea of risk stratification based on expected placebo response in that using symptom severity alone to develop a model that predicted the placebo response facilitated the building of a PI that moderated outcomes in CBT (vs. COMB and vs. MEDs). Prior research supports the notion that symptom severity is a reliable and strong indicator of prognosis. For example, Dinga et al. (2018) attempted to predict the 2-year prognosis of depression from a pool of 81 variables in 804 individuals. Despite a wide range of predictors that included biological variables (e.g., C-reactive protein (CRP), interleukin-6 (IL6), tumor necrosis factor-alpha) as well as clinical (e.g., co-morbid diagnoses, trauma) and psychosocial ones (e.g., demographics, personality traits), only baseline severity predicted future depression. Similarly, in the International Consortium to Predict PTSD (ICPP) sample ($N = 2473$) PTSD diagnostic status at follow-up was predicted by multiple variables including immediate/baseline post-trauma severity as well as female gender, education, and interpersonal nature of the trauma. However, the multiple variable model did not appear to be more accurate than a "severity-only" model (Shalev et al., 2019). An important clinical implication of these finding is that risk stratification may be less complicated for clinicians than has been assumed. It is a simple task to assess baseline severity of depression symptoms and this may be useful for the purposes of risk stratification.

Even if symptom severity proves to be the most useful variable for risk-stratification, depression symptoms are not interchangeable, some are better indicators of severity and impairment than others, and some may have different underlying contributors (Fried et al., 2016; Fried & Nesse, 2015; Zimmerman et al., 2006). Based on prior literature, we individually examined the effects of sleep problems and suicidality, but it is possible that considering all symptoms may offer advantages for the purposes of risk-stratification. Future research could also examine the interaction of severity with other variables like the duration of depression (Hollon et al., 2014; Nelson et al., 2013). Machine-learning algorithm like those that use regression trees are well-suited for modeling complex non-linear interactions, which has been recommended as a next step in treatment research (Lorenzo-Luaces & DeRubeis, 2018). Nevertheless, machine learning methods do not

always outperform other methods, which future research should take into account (Makridakis et al., 2018).

Adolescence is marked by at least two major developmental processes that likely affect treatment response; (1) the development of cognitive processes necessary for engagement in CBT (e.g., meta-cognition), and (2) greater emotional reactivity which may often override these cognitive abilities (Steinberg, 2005). The results of the TADS study, though they have generated critiques (Hollon et al., 2005) and critical tests of replicability (Davey et al., 2019), suggest that the combination of antidepressants, which may reduce emotional reactivity, and CBT, which may promote engagement in reappraisal and related cognitive processes (DeRubeis et al., 2008), may be the best treatment alternative for most adolescents with MDD. Our results confirm prior findings that symptom severity is a useful variable for risk-stratification, revealing a subgroup of patients who can benefit from monotherapy with CBT or antidepressants.

Acknowledgements

We wish to thank the patients of the TADS study for participating in the research as well as all investigators involved for generating the data and making it available.

Funding

This work was supported by American Psychological Foundation.

References

Akay, C. L., Meric-Bernstam, F., Hunt, K. K., Grubbs, E. G., Bedrosian, I., Tucker, S. L., Kuerer, H. M., Hoffman, K. E., Babiera, G. V., Strom, E. A., Buchholz, T. A., & Mittendorf, E. A. (2012). Evaluation of the MD Anderson Prognostic Index for local-regional recurrence after breast conserving therapy in patients receiving neoadjuvant chemotherapy. *Annals of Surgical Oncology*, *19*(3), 901–907. https://doi.org/10.1245/s10434-011-2006-7

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., Jüni, P., & Cuijpers, P. (2016). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *Focus*, *14*(2), 229–243. https://doi.org/10.1176/appi.focus.140201

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., Meyer, B., Huibers, M., Smit, F., van Straten, A., Warmerdam, L., Barkham, M., Bilich, L., Lovell, K., & Liu, E. T.-H. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ*, *346*(feb26 2), f540. https://doi.org/10.1136/bmj.f540

Chen, A. M., Meric-Bernstam, F., Hunt, K. K., Thames, H. D., Oswald, M. J., Outlaw, E. D., Strom, E. A., McNeese, M. D., Kuerer, H. M., Ross, M. I., Singletary, S. E., Ames, F. C., Feig, B. W., Sahin, A. A., Perkins, G. H., Schechter, N. R., Hortobagyi, G. N., & Buchholz, T. A. (2004). Breast conservation after neoadjuvant chemotherapy: The M.D. Anderson Cancer Center experience. *Journal of Clinical Oncology*, *22*(12), 2303–2312. https://doi.org/10.1200/JCO.2004.09.062

Cipriani, A., Zhou, X., Del Giovane, C., Hetrick, S. E., Qin, B., Whittington, C., Coghill, D., Zhang, Y., Hazell, P., Leucht, S., Cuijpers, P., Pu, J., Cohen, D., Ravindran, A. V., Liu, Y., Michael, K. D., Yang, L., Liu, L., & Xie, P. (2016). Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. *The Lancet*, *388*(10047), 881–890. https://doi.org/10.1016/S0140-6736(16)30385-3

Clayborne, Z. M., Varin, M., & Colman, I. (2019). Systematic review and meta-analysis: Adolescent depression and long-term psychosocial outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(1), 72–79. https://doi.org/10.1016/j.jaac.2018.07.896

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology*, *14*(1), 209–236. https://doi.org/10.1146/annurev-clinpsy-050817-084746

Curry, J., Rohde, P., Simons, A., Silva, S., Vitiello, B., Kratochvil, C., Reinecke, M., Feeny, N., Wells, K., Pathak, S., Weller, E., Rosenberg, D., Kennard, B., Robins, M., Ginsburg, G., & March, J. (2006). Predictors and moderators of acute outcome in the Treatment for Adolescents with Depression Study (TADS). *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(12), 1427–1439. https://doi.org/10.1097/01.chi.0000240838.78984.e2

Davey, C. G., Chanen, A. M., Hetrick, S. E., Cotton, S. M., Ratheesh, A., Amminger, G. P., & Harrison, B. J. (2019). The addition of fluoxetine to cognitive behavioural therapy for youth depression (YoDA-C): A randomised, double-blind, placebo-controlled, multicentre clinical trial. *The Lancet Psychiatry*, *6*, 735–744. https://doi.org/10.1016/S2215-0366(19)30215-9

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology*, *85*(9), 835–853. https://doi.org/10.1037/ccp0000231

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. https://doi.org/10.1016/j.brat.2016.02.003

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. https://doi.org/10.1371/journal.pone.0083875

DeRubeis, R. J., Siegle, G. J., & Hollon, S. D. (2008). Cognitive therapy versus medication for depression: Treatment outcomes and neural mechanisms. *Nature Reviews Neuroscience*, *9*(10), 788–796. https://doi.org/10.1038/nrn2345

Dinga, R., Marquand, A. F., Veltman, D. J., Beekman, A. T., Schoevers, R. A., van Hemert, A. M., Penninx, B. W. J. H., & Schmaal, L. (2018). Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: A machine learning approach. *Translational Psychiatry*, *8*(1), 241. https://doi.org/10.1038/s41398-018-0289-1

Driessen, E., Cuijpers, P., Hollon, S. D., & Dekker, J. J. (2010). Does pretreatment severity moderate the efficacy of

psychological treatment of adult outpatient depression? A meta-analysis. *Journal of Consulting and Clinical Psychology*, *78*(5), 668–680. https://doi.org/10.1037/a0020570

Eckshtain, D., Kuppens, S., Ugueto, A., Ng, M. Y., Vaughn-Coaxum, R., Corteselli, K., & Weisz, J. R. (2019). Meta-analysis: 13-year follow-up of psychotherapy effects on youth depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, Advanced online publication. https://doi.org/10.1016/j.jaac.2019.04.002

Essau, C. A., & Chang, W. C. (2009). Epidemiology, comorbidity, and course of adolescent depression. In C. A. Essau (Ed.), *Treatments for adolescent depression: Theory and practice* (pp. 3–25). Oxford University Press. https://doi.org/10.1093/med:psych/9780199226504.001.0001

Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Gallop, R., Amsterdam, J. D., & Hollon, S. D. (2008). Antidepressant medications *v.* cognitive therapy in people with depression with or without personality disorder. *British Journal of Psychiatry*, *192*(2), 124–129. https://doi.org/10.1192/bjp.bp.107.037234

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, *189*, 314–320. https://doi.org/10.1016/j.jad.2015.09.005

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, *13*(1), 72. https://doi.org/10.1186/s12916-015-0325-4

Garber, J., Weersing, V., Hollon, S., Porta, G., Clarke, G., Dickerson, J., Beardslee, W. R., Lynch, F. L., Gladstone, T. G., Shamseddeen, W., & Brent, D. A. (2018). Prevention of depression in at-risk adolescents: Moderators of long-term response. *Prevention Science*, *19*(S1), 6–15. https://doi.org/10.1007/s11121-015-0626-z

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Springer Series in Statistics). Springer.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, *41*(3), 924–936. https://doi.org/10.3758/BRM.41.3.924

Hollon, S. D., DeRubeis, R. J., Fawcett, J., Amsterdam, J. D., Shelton, R. C., Zajecka, J., Young, P. R., & Gallop, R. (2014). Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder. *JAMA Psychiatry*, *71*(10), 1157. https://doi.org/10.1001/jamapsychiatry.2014.1054

Hollon, S. D., Garber, J., & Shelton, R. C. (2005). Treatment of depression in adolescents with cognitive behavior therapy and medications: A commentary on the TADS project. *Cognitive and Behavioral Practice*, *12*(2), 149–155. https://doi.org/10.1016/S1077-7229(05)80019-7

Janssen, K. J., Vergouwe, Y., Donders, A. R. T., Harrell, F. E., Chen, Q., Grobbee, D. E., & Moons, K. G. (2009). Dealing with missing predictor values when applying clinical prediction models. *Clinical Chemistry*, *55*(5), 994–1001. https://doi.org/10.1373/clinchem.2008.115345

Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, *50*(2), 105–115. https://doi.org/10.1016/j.artmed.2010.05.002

Johnson, D. R., & Young, R. (2011). Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and Family*, *73*(5), 926–945. https://doi.org/10.1111/j.1741-3737.2011.00861.x

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. J. J. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980–988. https://doi.org/10.1097/00004583-199707000-00021

Kenward, M. G., & Molenberghs, G. (2009). Last observation carried forward: A crystal ball? *Journal of Biopharmaceutical Statistics*, *19*(5), 872–888. https://doi.org/10.1080/10543400903105406

Kessler, R. C. (2018). The potential of predictive analytics to provide clinical decision support in depression treatment planning. *Current Opinion in Psychiatry*, *31*(1), 32–39. https://doi.org/10.1097/YCO.0000000000000377

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. (2016). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, *26*, 22–36. https://doi.org/10.1017/s2045796016000020

Lachin, J. M. (2016). Fallacies of last observation carried forward analyses. *Clinical Trials*, *13*(2), 161–168. https://doi.org/10.1177/1740774515602688

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

Lewinsohn, P. M., Rohde, P., Seeley, J. R., Klein, D. N., & Gotlib, I. H. (2003). Psychosocial functioning of young adults who have experienced and recovered from major depressive disorder during adolescence. *Journal of Abnormal Psychology*, *112*(3), 353–363. https://doi.org/10.1037/0021-843X.112.3.353

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. (2018). Miles to go before we sleep: Advancing the understanding of psychotherapy by modeling complex processes. *Cognitive Therapy and Research*, *42*(2), 212–217. https://doi.org/10.1007/s10608-018-9893-x

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. https://doi.org/10.1016/j.jad.2017.02.010

Luedtke, A., Sadikova, E., & Kessler, R. C. (2018). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science*. https://doi.org/10.1177/2167702618815466

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way

forward. *International Journal of Forecasting*, *34*(4), 802–808. https://doi.org/10.1016/j.ijforecast.2018.06.001

March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J., Burns, B., Domino, M., McNulty, S., Vitiello, B., & Severe, J. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA*, *292*(7), 807–820. https://doi.org/10.1001/jama.292.7.807

Mayes, T. L., Bernstein, I. H., Haley, C. L., Kennard, B. D., & Emslie, G. J. (2010). Psychometric properties of the Children's Depression Rating Scale–Revised in adolescents. *Journal of Child and Adolescent Psychopharmacology*, *20*(6), 513–516. https://doi.org/10.1089/cap.2010.0063

Monroe, S., & Harkness, K. (2011). Recurrence in major depression: A conceptual analysis. *Psychological Review*, *118*(4), 655–674. https://doi.org/10.1037/a0025190

Monroe, S., & Harkness, K. (2012). Is depression a chronic mental illness? *Psychological Medicine*, *42*(5), 899–902. https://doi.org/10.1017/S0033291711002066

Moons, K. G., Donders, R. A., Stijnen, T., & Harrell Jr, F. E. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology*, *59*(10), 1092–1101. https://doi.org/10.1016/j.jclinepi.2006.01.009

Murray, C. J., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., Catherine, E. M., Kenji, S., Joshua, A. S., Safa, A., Victor, A., Jerry, A., Ilana, A., Rakesh, A., Stephanie, Y. A., Mohammed, K. A., Mohammad, A. A., Miriam, A., Ross, A. H., … Lopez, A. D. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), 2197–2223. https://doi.org/10.1016/S0140-6736(12)61689-4

National Institute for Health and Clinical Excellence (NICE). (2019). *Depression: Treatment and management of depression in adults. Clinical guideline 90*. www.nice.org.uk.

Nelson, J. C., Delucchi, K. L., & Schneider, L. S. (2013). Moderators of outcome in late-life depression: A patient-level meta-analysis. *American Journal of Psychiatry*, *170*(6), 651–659. https://doi.org/10.1176/appi.ajp.2012.12070927

Poznanski, E. O., & Mokros, H. B. (1996). *Children's depression rating scale, revised (CDRS-R)*. Western Psychological Services.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reynolds, W. M. (1987). *Professional manual for the reynolds adolescent depression scale*. Psychological Assessment Resources, Inc.

Reynolds, W. M. (2004). Reynolds adolescent depression scale – 2nd edition. In S. M. Hilsenroth (Ed.), *Comprehensive handbook of psychological assessment* (Vol. Volume 2: Personality assessment and psychopathology, pp. 224–236). John Wiley & Sons.

Rohde, P., Lewinsohn, P. M., Klein, D. N., Seeley, J. R., & Gau, J. M. (2013). Key characteristics of major depressive disorder occurring in childhood, adolescence, emerging adulthood, and adulthood. *Clinical Psychological Science*, *1*(1), 41–53. https://doi.org/10.1177/2167702612457599

Ruhé, H. G., Dekker, J. J., Peen, J., Holman, R., & De Jonghe, F. (2005). Clinical use of the Hamilton depression rating scale: Is increased efficiency possible? A post hoc comparison of Hamilton depression rating scale, Maier and Bech subscales, clinical global impression, and symptom checklist-90 scores. *Comprehensive Psychiatry*, *46*(6), 417–427. https://doi.org/10.1016/j.comppsych.2005.03.001

Rutherford, B. R., & Roose, S. P. (2013). A model of placebo response in antidepressant clinical trials. *American Journal of Psychiatry*, *170*(7), 723–733. https://doi.org/10.1176/appi.ajp.2012.12040474

Shalev, A. Y., Gevonden, M., Ratanatharathorn, A., Laska, E., van der Mei, W. F., Qi, W., Qi, W., Lowe, S., Lai, B. S., Bryant, R. A., Delahanty, D., Matsuoka, Y. J., Olff, M., Schnyder, U., Seedat, S., deRoon-Cassini, T. A., Kessler, R. C., & Koenen, K. C. (2019). Estimating the risk of PTSD in recent trauma survivors: Results of the International Consortium to Predict PTSD (ICPP). *World Psychiatry*, *18*(1), 77–87. https://doi.org/10.1002/wps.20608

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, *9*(2), 69–74. https://doi.org/10.1016/j.tics.2004.12.005

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

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*, *338*(jun29 1), b2393. https://doi.org/10.1136/bmj.b2393

TADS Team. (2003). Treatment for Adolescents with Depression Study (TADS): Rationale, design, and methods. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(5), 531–542. https://doi.org/10.1097/01.CHI.0000046839.90931.0D

TADS Team. (2005). The Treatment for Adolescents with Depression Study (TADS): Demographic and clinical characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(1), 28–40. https://doi.org/10.1097/01.chi.0000145807.09027.82

Trivedi, M. H., South, C., Jha, M. K., Rush, A. J., Cao, J., Kurian, B., Phillips, M., Pizzagalli, D. A., Trombello, J. M., Oquendo, M. A., Cooper, C., Dillon, D. G., Webb, C., Grannemann, B. D., Bruder, G., McGrath, P. J., Parsey, R., Weissman, M., & Fava, M. (2018). A novel strategy to identify placebo responders: Prediction index of clinical and biological markers in the EMBARC trial. *Psychotherapy and Psychosomatics*, *87*(5), 285–295. https://doi.org/10.1159/000491093

van Kuijk, S. M., Viechtbauer, W., Peeters, L. L., & Smits, L. (2016). Bias in regression coefficient estimates when assumptions for handling missing data are violated: A simulation study. *Epidemiology, Biostatistics and Public Health*, *13*, e11598-1–e11598-8.

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. J. J. (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

Webb, C. A., Trivedi, M. H., Cohen, Z. D., Dillon, D. G., Fournier, J. C., Goer, F., Fava, M., McGrath, P. J., Weissman, M., Parsey, R., Adams, P., Trombello, J. M., Cooper, C., Deldin, P., Oquendo, M. A., McInnis, M. G., Huys, Q., Bruder, G., Kurian, B. T., … Pizzagalli, D. A. (2018). Personalized prediction of antidepressant v. placebo response: Evidence from the EMBARC study. *Psychological Medicine*, *49*(7), 1118–1127. https://doi.org/10.1017/S0033291718001708

Weersing, V. R., Shamseddeen, W., Garber, J., Hollon, S. D., Clarke, G. N., Beardslee, W. R., Gladstone, T. R., Lynch, F. L., Porta, G., Iyengar, S., & Brent, D. A. (2016). Prevention of depression in at-risk adolescents: Predictors and moderators of acute effects. *Journal of the American Academy of Child &*

Adolescent Psychiatry, *55*(3), 219–226. https://doi.org/10.1016/j.jaac.2015.12.015

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., Jensen-Doss, A., Hawley, K. M., Krumholz Marchette, L. S., Chu, B. C., Weersing, V. R., & Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *American Psychologist*, *72*(2), 79–117. https://doi.org/10.1037/a0040360

Young, R., & Johnson, D. R. (2010). *Imputing the missing Y's: Implications for survey producers and survey users.* Paper presented at the Proceedings of the AAPOR conference abstracts.

Zhou, X., Hetrick, S. E., Cuijpers, P., Qin, B., Barth, J., Whittington, C. J., Cohen, D., Del Giovane, C., Liu, Y., Michael, K. D., Zhang, Y., Weisz, J. R., & Xie, P. (2015). Comparative efficacy and acceptability of psychotherapies for depression in children and adolescents: A systematic review and network meta-analysis. *World Psychiatry*, *14*(2), 207–222. https://doi.org/10.1002/wps.20217

Zimmerman, M., Chelminski, I., McGlinchey, J. B., & Young, D. (2006). Diagnosing major depressive disorder X: Can the utility of the DSM-IV symptom criteria be improved? *The Journal of Nervous and Mental Disease*, *194* (12), 893–897. https://doi.org/10.1097/01.nmd.0000248970.50265.34