Journal of Affective Disorders 279 (2021) 149–157

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Selecting the optimal treatment for a depressed individual: Clinical judgment or statistical prediction?

Suzanne C. van Bronswijk[a,*], Lotte H.J.M. Lemmens[a], Marcus J.H. Huibers[b,c], Frenk P.M.L. Peeters[a]

[a] Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200, MD Maastricht, the Netherlands
[b] Department of Clinical Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands
[c] Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104-6241, USA

ARTICLE INFO

Keywords:
Statistical prediction
Clinical judgment
Depression
Cognitive therapy
Interpersonal psychotherapy
Precision medicine

ABSTRACT

Background: Optimizing treatment selection is a way to enhance treatment success in major depressive disorder (MDD). In clinical practice, treatment selection heavily depends on clinical judgment. However, research has consistently shown that statistical prediction is as accurate - or more accurate - than predictions based on clinical judgment. In the context of new technological developments, the current aim was to compare the accuracy of clinical judgment versus statistical predictions in selecting cognitive therapy (CT) or interpersonal psychotherapy (IPT) for MDD.

Methods: Data came from a randomized trial comparing CT (n = 76) with IPT (n = 75) for MDD. Prior to randomization, therapists' recommendations were formulated during multidisciplinary staff meetings. Statistical predictions were based on Personalized Advantage Index models. Primary outcomes were post-treatment and 17-month follow-up depression severity. Secondary outcome was treatment dropout.

Results: Individuals receiving treatment according to their statistical prediction were less depressed at post-treatment and follow-up compared to those receiving their predicted non-indicated treatment. This difference was not found for recommended versus non-recommended treatments based on clinical judgment. Moreover, for individuals with an IPT recommendation by therapists, higher post-treatment and follow-up depression severity was found for those that actually received IPT compared to those that received CT. Recommendations based on statistical prediction and clinical judgment were not associated with differences in treatment dropout.

Limitations: Information on the clinical reasoning behind therapist recommendations was not collected, and statistical predictions were not externally validated.

Conclusions: Statistical prediction outperforms clinical judgment in treatment selection for MDD and has the potential to personalize treatment strategies.

1. Introduction

Despite the many options available, treatments for major depressive disorder (MDD) have modest effects on symptom reduction, with overall response rates around 50% (Papakostas and Fava, 2010). Although the average efficacy of various therapies is comparable (Cuijpers et al., 2011; Cuijpers et al., 2020a; Cuijpers et al., 2020b), individual treatment responses vary greatly (Simon and Perlis, 2010). As a result, depressed individuals may receive non-effective treatments before the right match is found (Rush et al., 2006). In addition, by offering treatments that do not work for a specific individual, risks of treatment dropout in MDD are high (Cooper and Conklin, 2015; Warden et al., 2014). To enhance treatment success, a straightforward solution is to optimize treatment selection for a given individual (what works best for whom?), moving beyond the *one size fits all* and *trial and error* approach towards *personalized* or *precision* strategies (Cohen and DeRubeis, 2018; DeRubeis, 2019).

In clinical psychology and psychiatry, attempts to match individuals to the most optimal treatment are everyday practice. Treatment selection heavily depends on clinical judgment of expected responses to available treatments. Clinical judgment is the result of informal or intuitive processes, and treatment recommendations are often formulated by an individual clinician or during multidisciplinary staff meetings. These recommendations are based on the clinicians' experience with

* Corresponding author.
E-mail address: suzanne.vanbronswijk@maastrichtuniversity.nl (S.C. van Bronswijk).

https://doi.org/10.1016/j.jad.2020.09.135
Received 23 March 2020; Received in revised form 25 July 2020; Accepted 27 September 2020
Available online 07 October 2020

similar patients, the clinicians' training background, the patient's preference and the clinicians' beliefs, theoretical perspective, and familiarity with empirical literature and (inter)national guidelines (Delgadillo et al., 2015; Lorenzo-Luaces et al., 2015).

Despite the potential flexibility of clinical judgment, human judgment is susceptible for many errors (Bell and Mellor, 2009; Garb, 2005; Grove et al., 2000; Richards et al., 2015). First of all, clinical judgment has shown to have low intra- and inter-rater reliability; clinical judgment is inconsistent for a given clinician over time, and agreement on the same case between clinicians is low (Bell and Mellor, 2009). Second, a source of error that contributes to the low inter-rater reliability is that clinicians use different theoretical frameworks that mediate their clinical judgement (Bell and Mellor, 2009). For example, a clinician with a behaviorist background will base clinical judgment on predominantly cognitive and behavioral factors, while an interpersonal oriented therapist will likely emphasize relational and attachment-style determinants. In addition, by basing predictions on causal theories, these theories need to be comprehensive and well-supported by evidence, and all variables that are relevant for these theories need to be available and measurable with accurate instruments (Grove and Meehl, 1996). No theory has yet met all of these conditions, for example illustrated by the lack of an exact understanding on working mechanisms of psychological treatments and biological determinants of psychopathology (Beijers et al., 2019; Cuijpers et al., 2019). Third, clinical judgment is prone to biases. As clinicians seldom rely on statistical rules only, they often apply cognitive heuristics. Although this might save time and effort, heuristics can lead to biases. In clinical judgment multiple types of biases are observed, including confirmation bias (bias through existing beliefs, expectations or hypotheses),anchoring bias (bias by initial information), availability bias (bias by easily available information) and representative biases (bias by assigning high probabilities to detailed prototypical combinations, i.e., the law of small numbers) (Bell and Mellor, 2009; Furnham and Boo, 2011; Nickerson, 1998; Richards et al., 2015). A fourth source of error is overreliance on unstructured clinical interviews that may lower the reliability and validity of clinical judgment. Explanations for this effect include confirmation bias as clinicians are more susceptible to focus on information that confirms their hypothesis and pay less attention to conflicting evidence (Bell and Mellor, 2009). A fifth source of error concerns the limited information processing abilities of humans. Even when biases are avoided, it is humanly impossible to differentially weight and combine large amounts of information (e.g., multiple conflicting predictors; Bell and Mellor, 2009; Meehl, 1986). A final source of error is thelack of adequate and systematic feedback that clinicians can use to change inaccurate clinical judgment behavior (Grove et al., 2000).

One way to overcome the errors of human judgment is to use statistical or actuarial predictions that are based on reproducible algorithms. The accuracy of these predictions, relative to predictions based on clinical judgment, has been a topic of debate since the mid-50's of the last century (Meehl, 1954, 1986). Meehl (1954) indicated that in 19 out of 20 studies statistical predictions were as accurate as or more accurate than predictions based on clinical judgment. Since then, research has consistently shown that statistical prediction works at least as well as predictions based on clinical judgment, with on average 10% to 13% more accuracy for statistical predictions (Ægisdóttir et al., 2006; Grove et al., 2000).

Despite the evidence that statistical prediction is at least as accurate as predictions based on clinical judgment, the use of statistical predictions in clinical practice is still limited (Bell and Mellor, 2009). Several potential reasons have been put forward to explain this phenomenon (Garb, 2000; Grove and Meehl, 1996; Katsikopoulos et al., 2008; Meehl, 1986). One explanation is the (expected) high demands of statistical prediction on the clinician's time. Another potential reason are ethical objections, for instance the viewpoint that ignoring a strong personal preference that conflicts with statistical prediction is *not right.* These ethical objections also come into play in the fear of dehumanization, that is the fear that statistical prediction reduces individuals to inanimate objects (numbers). Another explanation for the limited use of statistical predictions may be the clinician's clinging to theory, resulting in a reluctance to abandon theory-mediated predictions. In addition, the theoretical orientation of clinicians may be closely tied to their professional identity; poor performance of clinical judgement based on theoretical orientation may be perceived as a threat to this professional identity (*this is how I do it and that is who I am*). Fear of unemployment due to automated procedures is another potential reason for the limited use of statistical predictions in clinical practice. Finally, lack of education could be an important factor as well, by poor training in scientific techniques and human biases, and by role models involved in clinical training that ignore or disregard scientific reasoning and/or evidence.

However, the potential of novel statistical approaches (e.g., machine learning) and the availability of big data (e.g., electronic medical records and smartphone data) are leading us to a renewed attention for and rethink the importance of statistical predictions to personalize treatment recommendations (Delgadillo and Lutz, 2020; DeRubeis, 2019; Perna et al., 2018). In the context of treatment selection for MDD, this change is demonstrated by the development of multivariable models that promise to generate powerful predictions (Cohen and DeRubeis, 2018). One approach that combines information of multiple variables is the Personalized Advantage Index (PAI) that provides individual treatment recommendations by predicting the advantage of an indicated treatment over a non-indicated treatment (DeRubeis et al., 2014).

With the recent technological advances in mind, the aim of the current study was to compare the accuracy of clinical judgment versus statistical (PAI) predictions in the context of treatment selection of two frequently used psychotherapies (cognitive therapy [CT] and interpersonal psychotherapy [IPT]) for MDD. It was hypothesized that statistical prediction outperforms clinical judgment and that clinical judgment has only a very modest added value to statistical prediction in treatment selection for MDD.

2. Methods

2.1. Design and participants

Participants were recruited from the mood disorders unit of the Maastricht Outpatient Mental Health Centre (the Netherlands) and data was collected in the context of a randomized trial on the effectiveness of CT and IPT for MDD (for a detailed description: Lemmens et al., 2015; Lemmens et al., 2011). Participants had to have a primary diagnosis of MDD, internet access, an email address, and sufficient knowledge of the Dutch language. Exclusion criteria were bipolar or highly chronic depression (episode duration > 5 years), high acute suicide risk, concomitant pharmacological or psychological treatment, drugs and alcohol problems, and an IQ lower than 80. A total of 182 depressed outpatients were randomly assigned to CT (n = 76), IPT (n = 75), or a 2-month waiting-list control condition followed by treatment of participants' choice (n = 31). In this study we focused on data of individuals randomized to CT and IPT. Participants received 16-20 weekly sessions of 45 minutes (m = 17.0, SD = 2.9), using CT and IPT protocols (Beck et al., 1979; Klerman et al., 1984). Treatments were provided by ten therapists (licensed psychologist, psychotherapists, and psychiatrists) with on average 9.1 years of experience (SD = 5.4, range 4-21 years). There was no significant effect of individual therapists on treatment outcomes (Lemmens et al., 2015). Participating therapists delivered either CT or IPT to avoid contamination of treatment conditions. Quality of treatment was rated good to excellent with significant differences in therapy-specific behavior between treatments. Written informed consent was obtained and the study was approved by the Medical Ethics Committee of Maastricht University. The study is

registered at the Netherlands Trial Register, part of the Dutch Cochrane Centre (ISRCTN67561918).

2.2. Outcome variables

Primary outcome was depression severity, measured with the Beck Depression Inventory II (BDI-II) during treatment (month 0 – 7) and follow-up (month 7- 24, Beck et al., 1996). For the current study, we focused on post-treatment BDI-II scores (month 7) and follow-up BDI-II scores (month 7, 8, 9, 10, 11, 12 and 24, aggregated into an Area Under the Curve, see van Bronswijk et al., 2019). Treatment dropout (discontinuation of the treatment intervention by participants) was examined as a secondary outcome.

2.3. Therapists' treatment recommendations

Therapists' treatment recommendations (CT, IPT or no specific recommendation) were based on case presentations by clinicians who carried out the diagnostic work-up during regular multidisciplinary team meetings at the mood disorders unit and were formulated after group-discussion. Team meetings were attended by a total of 4-5 clinicians (psychologists, psychotherapists and psychiatrists) that had allegiance to both CT and IPT. The specific composition of the attendees depended on their involvement in the diagnostic work-up procedure and on scheduling matters. Part of the attendees of the team meetings were involved as therapists in the study as well. Recommendations were formulated before randomization of the participants. After randomization, participants were coded as having received the recommended or non-recommended treatment through random allocation (CT or IPT). Treatment recommendations were administered for a subset of the participants (n = 110, CT = 55, IPT = 52, no preference for CT or IPT = 3). For the current analyses we limited our sample to those that had a recorded treatment recommendation for either CT or IPT (n = 107, CT = 55, IPT = 52).

2.4. Statistical predictions

Statistical treatment predictions were based on algorithms generating the PAI, a measure of the predicted advantage of the indicated therapy (CT or IPT) as compared to the other non-indicated therapy. In this study, two types of PAI scores were used: one that was focused on post-treatment BDI-II predictions(post-treatment PAI, Huibers et al., 2015)and one that was based on follow-up BDI-II predictions(long-term PAI, van Bronswijk et al., 2019).Variables included in the PAI-models were *predictors* (prognostic factors that predict outcome irrespective of the received treatment type) and *moderators* (prescriptive factors that predict outcome as a function of treatment type) measured before treatment and randomization. For the post-treatment PAI, predictors and moderators were identified using a modified domain approach that involves a series of linear regression models for different predictor domains(e.g., history of illness or demographics; Fournier et al., 2008). Predictors and moderators were then combined in a linear regression model from which individual predictions werecomputed using a leave-one-out cross-validation approach (Picard and Cook, 1984). For each individual, a separate prediction for CT and for IPT was computed. The difference between these two predictions was a positive or negative score indicating the PAI-recommended treatment: CT or IPT (DeRubeis et al., 2014). For the long-term treatment PAI model, a two-step machine learning approach was used to identify predictors and moderators: the application of a random forest algorithm (Garge et al., 2013),followed by a backward elimination approach using multiple bootstrapped samples (Austin and Tu, 2004; Rizopoulos and Rizopoulos, 2009). The selected variables were combined in a regression model, and long-term outcomes were predicted for CT and IPT using a fivefold cross-validation (Picard and Cook, 1984)for each individual, and with these predictions, individual PAI scores were computed (DeRubeis et al., 2014). The post-treatment PAI scores were calculated for individuals that had a post-treatment BDI-II score (n = 134, CT = 69, IPT = 65; Huibers et al., 2015), while for the long-term PAI all individuals were included by applying an imputation technique for missing data based on a non-parametric random forest approach (Stekhoven and Bühlmann, 2012; van Bronswijk et al., 2019; n = 151, CT = 76, IPT = 75). The two PAI models were developed after study termination, and therapists were therefore blind to these estimates. Participants were coded as having received a PAI-indicated or non-indicated treatment according to the post-treatment and long-term PAI models. More detailed information about the development of the PAI models can be found elsewhere (Huibers et al., 2015; van Bronswijk et al., 2019).

2.5. Statistical analyses

First, pre-treatment characteristics were compared between individuals that did not have a recorded therapists' recommendation (n = 41) and individuals that had a recorded therapists' recommendation for CT or IPT and were included in the analyses (n = 107). For these comparisons, t-tests and $\chi 2$-tests were applied for continuous and categorical variables respectively. Second, pre-treatment characteristics were examined between the participants that had a CT versus an IPT recommendation according to clinical judgment using t-tests and $\chi 2$-tests where appropriate. Significant differences in pre-treatment characteristics between CT versus IPT recommendations ($p \leq 0.10$ in the t-tests or $\chi 2$-tests) were further examined using a logistic regression model. In this logistic regression model, therapist preference was the dependent variable and the pre-treatment variables were the independent variables. Then, the overlap between these pre-treatment variables with the pre-treatment variables that are part of the algorithms of the PAI models was evaluated. Third, the percentage of agreement and kappa coefficients were calculated to examine the level of agreement between PAI recommendations and therapists' recommendations. Fourth, average post-treatment and follow-up depression severity (BDI-II scores) of patients receiving a therapists' recommended versus a non-recommended treatment were compared using t-tests. The results of these t-tests were then compared to the results of previously conducted t-tests on post-treatment and follow-up depression severity of individuals receiving PAI indicated versus PAI non-indicated treatment (Huibers et al., 2015; van Bronswijk et al., 2019). The comparisons between therapists' recommendations and post-treatment PAI recommendations were applied to a subset of the individuals (n = 95, CT = 51, IPT = 44), since 12 individuals with a therapist recommendation did not have a post-treatment PAI recommendation. For the comparisons between therapists' recommendations and long-term PAI recommendations, all individuals with a therapist recommendation had a long-term PAI recommendation and were therefore included in the analyses (n = 107, CT = 55, IPT = 52). Then, for the purpose of comparison between clinical judgment and statistical predictions, three separate regression models were constructed for both post-treatment BDI-II scores and long-term BDI-II scores as dependent variables. For each model, different predictors were included: model 1 included the statistical PAI recommendation for the optimal treatment (the statistical model), model 2 included the therapists' recommendation for the optimal treatment (the clinical judgment model) and model 3 included both the statistical PAI recommendation and therapist recommendation (the combined model). All six models were corrected for depression severity at baseline since there was a small but non-significant difference between CT and IPT at baseline (Lemmens et al., 2015). Finally, treatment dropout rates were compared between individuals receiving a therapists' recommended versus a non-recommended treatment, and between individuals receiving a PAI indicated versus a PAI non-indicated treatment using $\chi 2$-tests.

3. Results

3.1. Pre-treatment characteristics of recorded vs. non recorded therapists' recommendation

First, pre-treatment characteristics were compared between individuals that did not have a recorded therapists' recommendation (n = 41) with those that had a recorded therapists' recommendation for CT or IPT and were included in the analyses (n = 107). Results indicated that individuals with a recorded therapists' recommendation were younger (m = 39.6, SD = 12.3 vs. m = 45.7, SD =10.5, t = -2.80, p = 0.01), had lower functional impairment (Work and Social Adjustment Scale; m = 23.6, SD = 7.6 vs. m = 20.4, SD = 6.7, t = 2.38, p = 0.02), had less cognitive complaints (Brief Symptom Inventory subscale; m = 11.9, SD = 4.6 vs. m = 9.6, SD = 5.61, t = 2.57, p = 0.01), and hand were more likely to have a recurrent depression (63.4% vs. 44.9%, $\chi 2$ (1, n = 107) = 4.08, p = 0.04).

3.2. Pre-treatment characteristics associated with therapists' and PAI-based recommendations

In Table 1, comparisons between pre-treatment characteristics between the participants that had a CT versus an IPT recommendation by therapists are shown. The proportion of individuals with a comorbid personality disorder was higher for individuals with a CT recommendation versus those with an IPT recommendation according to clinical judgment (χ^2 (1)=3.60, p=0.058). This was not in accordance with the pre-treatment characteristics of the PAI models: personality disorder status appeared only as a predictor (not a moderator) for the post-treatment PAI model, indicating that this variable predicted outcome irrespective of the received treatment (Huibers et al., 2015). In addition, personality disorder status was not part of the long-term outcome PAI algorithm. The level of pre-treatment somatic complaints was significantly higher for individuals with a CT recommendation compared to those with an IPT recommendation by therapists (t = 2.05, df = 105, p = 0.043). This was in accordance with the post-treatment PAI model. In the post-treatment PAI model, somatic complaints were related to a better response to CT as compared to IPT

Table 1
Predictors and moderators of the post-treatment and long-term PAI models and comparisons of pre-treatment characteristics between individuals with CT versus IPT therapist recommendations.

Post-treatment PAI model		
Predictors	Baseline depression severity (BDI-II), gender, employment status, anxiety symptoms (BSI), personality disorder (SCID-II), quality of life utility score (EQ-5D)	
Moderators	Somatic complaints (BSI), cognitive problems (BSI), paranoid symptoms (BSI), self-sacrificing (IIP), attributional style on achievement (ASQ), number of life events in past year	
Long-term PAI model		
Predictors	Parental alcohol abuse	
Moderators	Number of life events in past year, number of childhood trauma events	
Therapist recommendations		
	CT recommendation	**IPT recommendation**
Demographics		
Age, years, mean (*SD*)	39.0 (12.7)	40.2 (12.1)
Female, *n* (%)	36 (67.9%)	34 (63.0)
Partner, *n* (%)	33 (62.3%)	32 (59.3%)
Employed, *n* (%)	34 (64.2%)	33 (61.1%)
Education level		
- Low, *n* (%)	- 9 (17.0%)	- 13 (24.1%)
- Intermediate, *n* (%)	-35 (66.0%)	- 31 (57.4%)
- High, *n* (%)	- 9 (17.0%)	- 10 (18.5%)
Clinical features		
BDI-II baseline score, mean (*SD*)	30.8 (6.6)	29.1 (9.4)
Recurrent depression,*n* (%)	26 (49.1%)	22 (40.7%)
Comorbid personality disorders (SCID-II), *n* (%)	21 (40.4)*	12 (23.1)*
Other co-morbid disorders (SCID-I), *n* (%)	24 (45.3)	31 (57.4)
Psychological distress (BSI), mean (*SD*)	73.7 (23.4)	67.6 (31.1)
Anxiety symptoms (BSI), mean (*SD*)	8.6 (4.6)	8.1 (4.5)
Somatic complaints (BSI), mean (*SD*)	8.8 (5.3)*	6.9 (4.2)*
Cognitive problems (BSI), mean (*SD*)	12.2 (4.5)	11.7 (4.8)
Paranoid symptoms (BSI), mean (*SD*)	7.2 (3.3)	6.9 (4.9)
Attributional style on achievement (ASQ), mean (*SD*)	5.1 (0.8)	4.8 (1.0)
Self-sacrificing (IIP), mean (*SD*)	13.5 (5.2)	14.6 (6.2)
Functionality		
WSAS, mean (*SD*)	24.7 (6.7)	22.5 (8.3)
EQ-5D, mean (*SD*)	0 .6 (0.2)	0 .6 (0.2)
RAND-36, mean (*SD*)	14.1 (2.8)	14.2 (3.1)
Life and family history		
Number of life events last year, mean (*SD*)	1.5 (1.4)*	2.0 (1.4)*
Number of childhood trauma events, mean (*SD*)	0.9 (1.1)	0.9 (1.2)
Parental alcohol abuse, yes (%)	12 (22.6)	6 (11.1)

Note: BDI-II, Beck Depression Inventory, second edition; BSI, Brief Symptom Inventory; SCID-II, Structured Clinical Interview for DSM-IV Axis II disorders; EQ-5D, EuroQol 5D; IIP, Inventory of Interpersonal Problems; ASQ, Attributional Style Questionnaire; SCID-I Structured Clinical Interview for DSM-IV Axis I disorders; WSAS, Work and Social Adjustment Scale
Predictors: prognostic factors that predict outcome irrespective of the received treatment type
Moderators: prescriptive factors that predict outcome as a function of treatment type
* $p < 0.10$

(Huibers et al., 2015). The number of life events in the past year was significantly higher in individuals with an IPT recommendation versus a CT recommendation based on clinical judgment (t = -2.18, df = 105, p = 0.03). This is in contrast to the post-treatment PAI model and the long-term PAI model. In both models, more life events were associated with a better response in CT as compared to IPT (Huibers et al., 2015; van Bronswijk et al., 2019). The presence of a comorbid personality disorder, the level of pre-treatment somatic complaints and the number of life events were then combined in one logistic regression model with therapist recommendations as the outcome variable. Results of the model indicated that the presence of a comorbid personality disorder was associated with a CT recommendation (β = -.75, p = 0.104; trend significant). In addition, the level of pre-treatment somatic complaints was significantly associated with a CT recommendation (β = -0.10, p = 0.025), while the number of life events in the past year was significantly associated with an IPT recommendation (β = 0.33, p = 0.041).

3.3. Comparisons between therapists' and PAI-based recommendations

There was no agreement between therapists' recommendations and post-treatment PAI recommendations (47.4% agreement, kappa coefficient = -0.07). In addition, there was no agreement between therapists' recommendations and long-term PAI recommendations (43.0% agreement, kappa coefficient = -0.14). As illustrated in Fig. 1a, actual post-treatment BDI-II scores were (significantly) lower for individuals randomized to their PAI-indicated treatment as compared to individuals allocated to their PAI non-indicated treatment for both the CT and IPT condition (for a more detailed description, see Huibers et al., 2015). The opposite pattern was found for the treatment recommendations based on clinical judgment: the post-treatment BDI-II scores were *higher* for individuals who received the therapist recommended treatment as compared to those who received their non-recommended treatment (recommended treatment: n = 46, mean = 17.1, non-recommended treatment: n = 49, mean = 12.6, t = -1.92, df = 93, p = 0.058, d = -0.39 (CI = -0.80 to 0.01)). In the CT condition this difference was not significant (recommended treatment: n = 23, mean = 15.9, non-recommended treatment: n = 21, mean = 13.8, t = 0.67, df = 42, p = 0.508, d = 0.20 (CI = -0.39 to 0.79), Fig. 1b). In IPT, this difference was borderline significant (recommended treatment: n = 23, mean = 18.3, non-recommended treatment: n = 28, mean = 11.7, t = -1.89, df = 49, p = 0.064, d = -0.53 (CI = -1.09 to 0.03), Fig. 1b). As shown in Fig. 2a, follow-up BDI-II scores were significantly lower for individuals allocated to their PAI-indicated treatment as compared to individuals allocated to their PAI non-indicated treatment in the CT condition, but not in the IPT condition (for a more detailed description, see van Bronswijk et al., 2019). Again, for the recommendation based on clinical judgment, opposite patterns were found with non-significant higher BDI-II follow-up scores for individuals who received their recommended treatment as compared to those that got their non-recommended treatment (recommended treatment: n = 51, mean = 16.3, non-recommended treatment:n = 56, mean = 14.3, t = -0.95, df = 105, p = 0.347, d = -0.18 (CI = -0.56 to 0.20)). This difference was not significant in the CT condition (recommended treatment: n = 26, mean = 14.2, non-recommended treatment:n = 27, mean = 15.8, t = -0.57, df = 51, p = 0.574, d = -0.16 (CI = -0.69 to 0.38), Fig. 2b). In IPT, individuals with an IPT recommendation had borderline significantly higher follow-up BDI-II scores compared to those with a CT recommendation (recommended treatment: n = 25, mean = 18.5, non-recommended treatment: n = 29, mean = 12.8, t = -1.77, df = 55, p = 0.082, d = -0.48 (CI = -1.02 to 0.06), Fig. 2b).

Model comparisons of statistical predictions versus therapists' recommendations are shown in Table 2. The statistical models indicate that the PAI recommendation (PAI-indicated vs. PAI non-indicated treatment) was (borderline) significantly associated with post-treatment and follow-up BDI-II scores, with lower depression severity for individuals receiving a PAI-indicated treatment. In contrast, the therapists' recommendations in the clinical judgment model were not significantly related to post-treatment and follow-up depression severity. When both the PAI recommendations and the therapists' recommendations variables were included in the combined model, only the statistical prediction was associated with post-treatment and follow-up BDI-II scores. In addition, the therapists' recommendation failed to add predictive value to the models in terms of explaining the variance (adjusted R^2).

As shown in Table 3, the proportion of treatment dropouts was not significantly different in individuals randomized to their PAI-indicated treatment as compared to individuals allocated to their PAI non-indicated treatment in both the post-treatment (χ^2 (1)= 0.12, p=0.730) and long-term PAI model (χ^2 (1)= 1.51, p=0.22). In addition, there were no differences in dropout rates between individuals that received their recommended versus their non-recommended treatment according to clinical judgment (χ^2 (1)= 0.25, p=0.62).

4. Discussion

The current study compared the accuracy of clinical recommendations versus statistical PAI predictions in the context of treatment selection of CT versus IPT for MDD. There was limited overlap between pre-treatment variables associated with therapists' recommendations and pre-treatment variables included in the PAI algorithms. Overall, therapists were more likely to recommend CT to individuals that had a comorbid personality disorder (trend significant) and a higher level of somatic complaints, while an IPT recommendation was more common in individuals with a recent life event. For somatic complaints, the IPT recommendation was in accordance with the post-treatment PAI model. However, comorbid personality disorder was not a moderator in the PAI models and the presence of current life events was, in contrast to the therapist recommendation, identified as a moderator for worse outcomes in IPT in both PAI models. As we reported before, participants that received their indicated treatment according to their PAI score had lower post-treatment and follow-up depression severity as compared to those that received their PAI non-indicated treatment (for follow-up depression severity, this was only the case for the CT condition, Huibers et al., 2015; van Bronswijk et al., 2019). Interestingly, a reverse pattern was found for the therapists' IPT recommendations: participants that received the therapist recommended treatment had higher post-treatment and follow-up depression severity as compared to those receiving their non-recommended treatment. For individuals with a therapists' recommendation for CT, post-treatment and follow-up depression severity did not differ between those who received the recommended versus the non- recommended treatment. In addition, model comparisons showed a significant association between statistical prediction and depression severity, both at post-treatment and follow-up, and demonstrated that therapists' recommendations had no added value to statistical prediction for treatment selection in MDD. Finally, both statistical predictions and therapist recommendations were not associated with treatment dropout; there was no significant difference in treatment dropout rates between individuals assigned to a PAI-indicated or recommended treatment versus those randomized to a PAI non-indicated or non-recommended treatment.

In general, the results indicate that individuals that received a CT recommendation by therapists had more internal difficulties (personality disorder and somatic complaints), whilethe individuals that were given an IPT recommendation suffered more from external stressors. There are several possible explanations for these different patterns of treatment recommendation. The finding that a comorbid personality disorder was associated with a CT recommendation by therapists could be explained by awareness of therapists of (inconsistent) research findings of CT versus IPT head-to-head studies that suggest that personality disorder features are treatment moderators (Barber and Muenz, 1996; Carter et al., 2011; Joyce et al., 2007; Ryder et al., 2010). Other explanations could be that therapists perceived the more

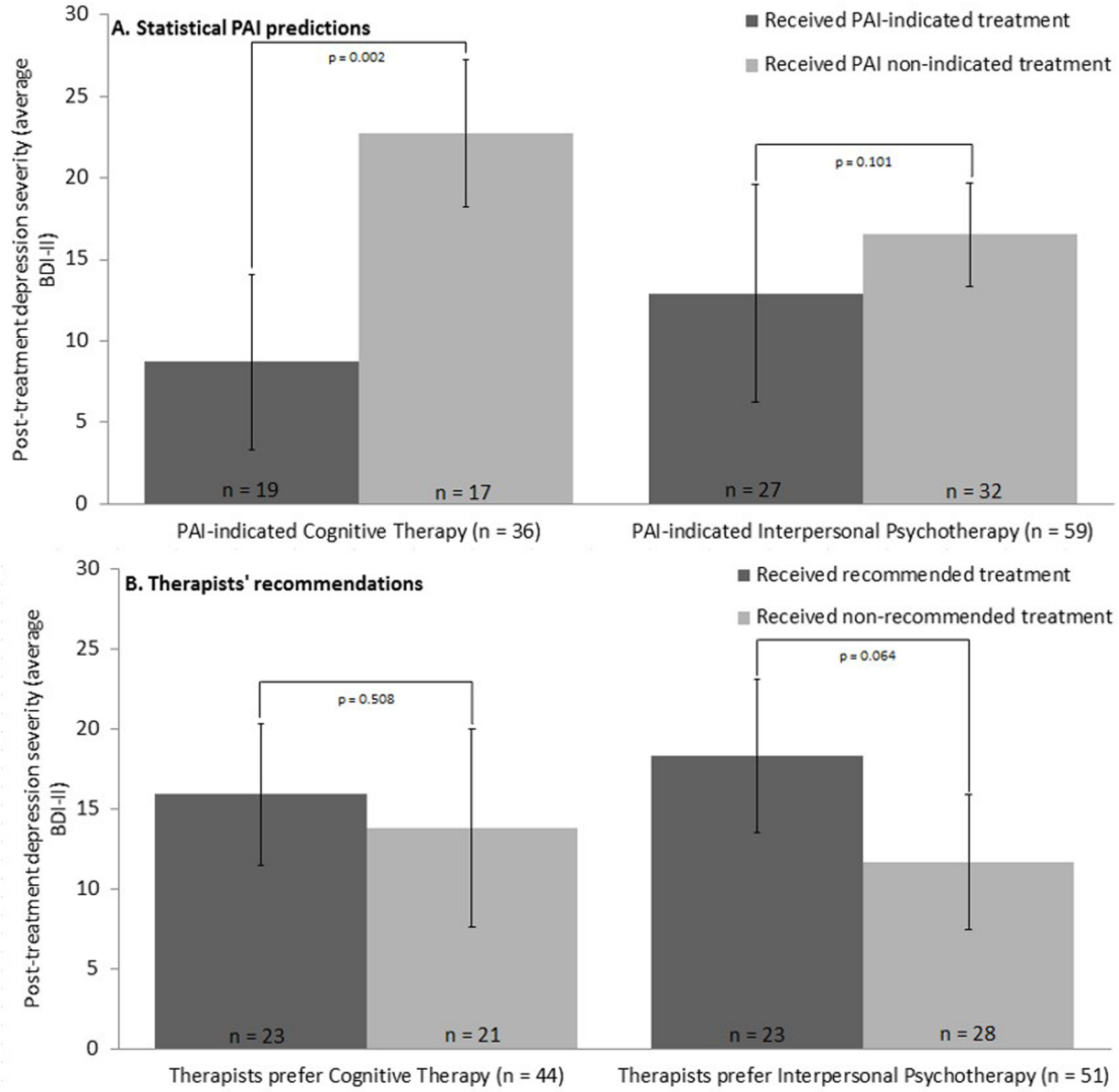

Fig. 1. Comparison of the observed post-treatment BDI-II scores for individuals randomly assigned to their PAI-indicated optimal treatment versus their PAI-indicated non-optimal treatment (A) and for individuals randomly assigned to their recommended treatment versus non-recommended treatment by the therapists (B).

*Note: Based on the availability of therapists' recommendations and statistical predictions, a subset of 95 individuals was used to compare post-treatment outcomes (CT = 51, IPT = 44). PAI predictions of individuals that completed the pre- and posttreatment assessment of the BDI-II (n = 134) can be found elsewhere (Huibers *et al.*, 2015).

structured nature of CT a better match for individuals with a comorbid PD or that it was thought to be too difficult to find a specific interpersonal focus for IPT since complex interpersonal problems are common in depressed individuals with a comorbid personality disorder (Markowitz et al., 2007). The finding that a higher level of somatic complaints was associated with a CT recommendation by therapists, could be explained by knowledge of therapists on the role of CT in treatment of somatization (Kroenke, 2007). Another explanation could be that therapists thought that the focus on exposure in CT through *in vivo* homework assignments was a better match for individuals with physical problems. The finding that life events in the past year was associated with an IPT recommendation was in contrast to both PAI models, however not surprising given that three of the four treatment foci of IPT relate to recent life events (complicated bereavement, role dispute, and role transition, Klerman et al., 1984). An explanation for the IPT recommendation could therefore be that therapists aim to target (multiple) recent life events by assigning individuals to IPT.

The results of this study confirmed the finding that statistical prediction is at least as accurate as or more accurate than clinical recommendation (Ægisdóttir et al., 2006; Grove et al., 2000). The results are also in line with a recent study that demonstrated that prognostic information based on machine learning techniques outperforms clinical judgment in predicting social functioning in individuals with recent-onset MDD and psychosis (Koutsouleris et al., 2018). In addition to the finding that statistical prediction was more accurate then clinical judgment, our study indicated that participants receiving IPT as recommended by the therapists had higher post-treatment and follow-up depression severity compared to those receiving their non-recommended treatment. Although this finding was only trend significant and needs to be replicated, it points to potential iatrogenic effects of

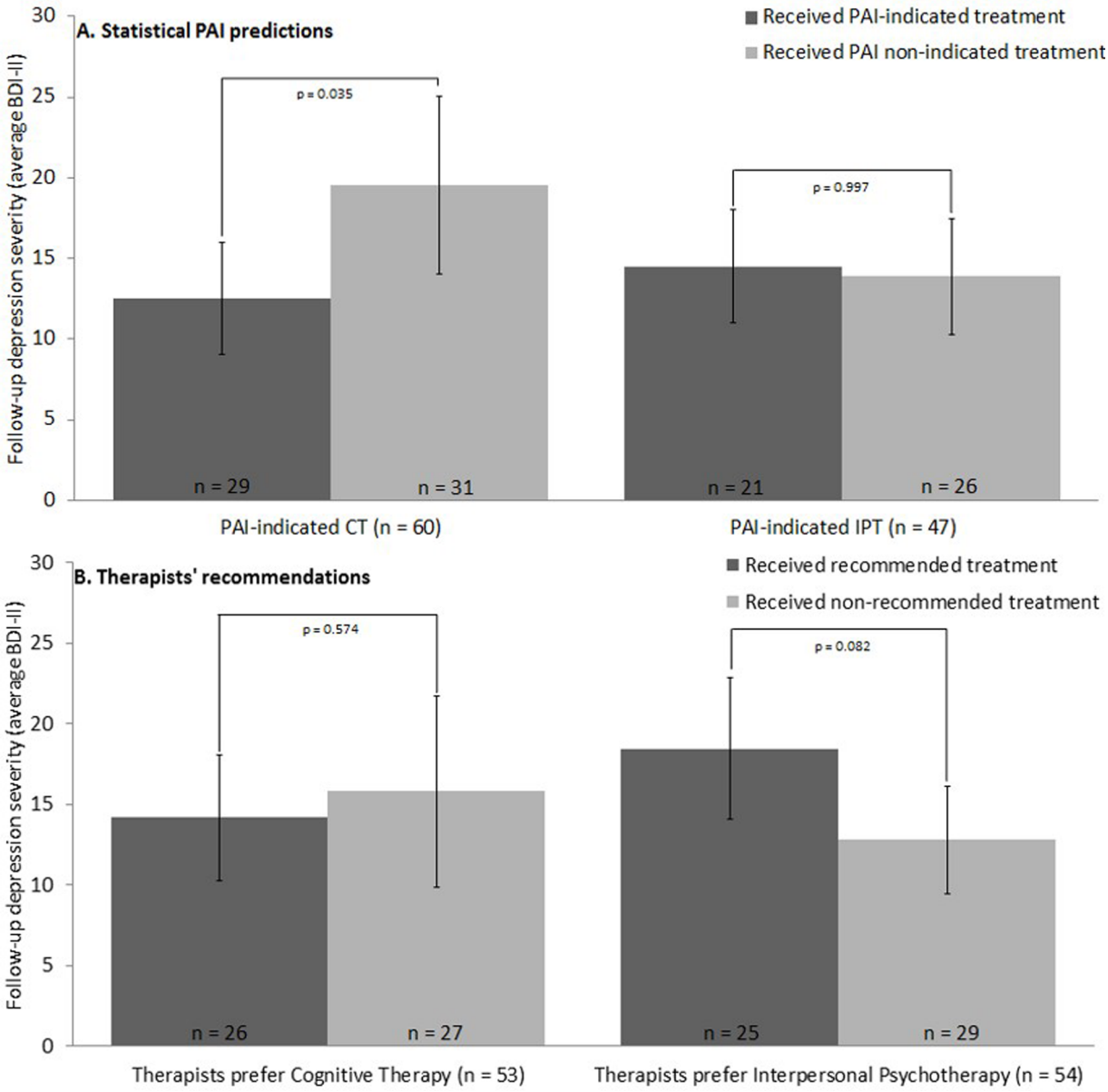

Fig. 2. Comparison of the observed follow-up BDI-II scores for individuals randomly assigned to their PAI-indicated optimal treatment versus their PAI-indicated non-optimal treatment (A) and for individuals randomly assigned to their recommended treatment versus non-recommended treatment by the therapists (B).
*Note: Based on the availability of therapists' recommendations and statistical predictions, a subset of 107 individuals was used to compare follow-up outcomes (CT = 55, IPT = 52). PAI predictions of the total sample (n = 151) can be found elsewhere (van Bronswijk *et al.*, 2019).

current clinical judgement in treatment selection for MDD. Based on previous studies, a few elements could explain the therapists' low predictive accuracy in the process of treatment selection for MDD. First, therapists involved in the clinical judgment came from the same setting as the study participants, which is associated with less accuracy compared to clinicians that make predictions for another setting (Ægisdóttir et al., 2006). Second, clinicians based their recommendations mainly on a clinical interview, which is known to negatively affect the prediction of outcome (Grove et al., 2000). Third, the statistical prediction was based on an algorithm, which has been proven to be more accurate then logically constructed rules (Ægisdóttir et al., 2006). Finally, the statistical prediction is based on state-of-the-art machine learning methods including cross-validation, and model-based recursive partitioning (Huibers et al., 2015; van Bronswijk et al., 2019).

The final finding of this study was that both statistical prediction and therapist recommendations were not associated with different rates of treatment dropout. The absence of an association between statistical prediction and treatment dropout was not surprising, since treatment dropout is potentially related to other variables than treatment outcome (Swift et al., 2017), and involves different types of PAI models (Keefe et al., 2018; Zilcha-Mano et al., 2016). However, this finding is in contrast to what one might expect from clinical judgment, since treatment allocation in clinical practice is not only based on clinical predictions of treatment outcome, but also on clinical predictions of treatment tolerance and engagement.

This study has some limitations. First, therapists' treatment recommendations were based on case presentations by the interviewers and formulated after discussion during regular team meetings. Although this setting is representative for clinical practice, one could argue that the clinicians had access to less data compared to the variables included in the statistical algorithms. However, previous research has shown that increasing the amount of information decreases the clinicians' judgment accuracy, and that access to all variables included in statistical formulas does not improve the clinicians' accuracy

Table 2
Comparison of models with different predictors: statistical prediction (model 1, "statistical"), therapists' recommendation (model 2, "clinical") and both statistical prediction and therapists' recommendations (model 3, "combined").

Outcome: post-treatment BDI-II	Statistical model			Clinical judgment model			Combined model		
Variable	Coef.	*SE*	*p*	Coef.	*SE*	*p*	Coef.	*SE*	*p*
Intercept	17.49	1.27	<0.0001	13.20	1.52	<0.0001	16.31	1.80	<0.0001
BDI-II baseline	5.40	0.91	<0.0001	5.24	1.22	<0.0001	4.64	1.19	<0.0001
Statistical prediction (optimal or non-optimal)	-5.38	1.82	0.004				-6.23	2.12	0.004
Therapists' recommendation (optimal or non-optimal)				3.04	2.20	0.171	2.88	2.12	0.178
Adjusted R^2	0.25			0.18			0.24		
Outcome: follow-up BDI-II	**Statistical model**			**Clinical judgment model**			**Combined model**		
Variable	Coef.	*SE*	*p*	Coef.	*SE*	*p*	Coef.	*SE*	*p*
Intercept	17.48	1.21	<0.0001	15.01	1.36	<0.0001	17.22	1.71	<0.0001
BDI-II baseline	5.32	0.87	<0.0001	5.28	1.10	<0.0001	5.46	1.09	<0.0001
Statistical prediction (optimal or non-optimal)	-3.27	1.73	0.061				-4.07	1.96	0.040
Therapists' recommendation (optimal or non-optimal)				0.26	2.00	0.896	-0.38	1.99	0.847
AdjustedR^2	0.20			0.17			0.20		

Note: SE, standard error; BDI-II, Beck Depression Inventory, second edition

Table 3
Comparison of treatment dropout rates for individuals randomly assigned to their PAI-indicated optimal treatment versus their PAI-indicated non-optimal treatment, and for individuals randomly assigned to their recommended treatment versus non-recommended treatment by the therapists

	CT recommendation Received CT	Received IPT	IPT recommendation Received IPT	Received CT
Post-treatment PAI model				
Treatment dropouts, *n* (%)	6 (31.6)	5 (29.4)	3 (11.1)	6 (18.8)
Long-term PAI model				
Treatment dropouts, *n* (%)	6 (20.1)	7 (22.6)	3 (14.3)	9 (34.6)
Therapist recommendations				
Treatment dropouts, *n* (%)	6 (23.1)	3 (11.1)	7 (28.0)	9 (31.0)

*Note: Based on the availability of therapists' recommendations and statistical predictions, a subset of 95 individuals was used to compare dropout rates for the post-treatment PAI model (CT = 51, IPT = 44), and a subset of 107 individuals was used for the long-term PAI model and the therapist recommendations (CT = 55, IPT = 52).

(Ægisdóttir et al., 2006). Second, since treatment recommendations were formulated during regular interdisciplinary team meetings, the level of expertise of the attendees varied, which could have influenced the accuracy of the clinical predictions (Ægisdóttir et al., 2006). However, this remains unclear as there is conflicting evidence that the level of training, general experience, and task-related experience has impact on the correctness of clinical judgment (Grove et al., 2000). Third, although we link pre-treatment variables to the therapist recommendations, information on the clinical reasoning behind this recommendation was not collected. In addition, detailed information on the therapists involved in the team meetings was not available either. Fourth, for a substantial subset of all study participants no therapists' recommendation was available. This subset had significantly different pre-treatment characteristics (i.e., younger age, less cognitive problems, less functioning impairment and a higher percentage of recurrent depression) compared to the participants with a recorded treatment recommendation. However, these missing assessments were related to time constraints that are common in daily clinical practice, and are therefore more likely to be random (instead of related to the specific participant characteristics). Finally, although cross-validation techniques were applied to minimize the riskof overfitting, the PAI models were not tested in an independent dataset, which is a limitation of most existing prediction models in mental health (Cohen and DeRubeis, 2018; DeRubeis, 2019). Without external validation and prospective testing, it remains unknown how well these models generalize to clinical practice.

The implementation of statistical prediction for treatment selection in MDD has great potential, particularly in the context of new statistical approaches and data availability (Cohen and DeRubeis, 2018; Perna et al., 2018). Moreover, one might argue that it is unethical to not implement these new approaches (Dawes, 2005). However, as history shows us, scientific findings on superiority of statistical prediction relative to clinical judgment are not sufficient to change mental health care practice. How can we advance the use of these empirical supported predictions in routine clinical practice? One suggestion is by changing clinical training. Education on effective clinical-decision making should be a key part of clinical training, including adequate statistics, probability theory, and education on human judgment errors (Ægisdóttir et al., 2006). Another suggestion is to provide feedback on treatment decision regularly and systematically (Knaup et al., 2009). Clinicians with proper training should collect and evaluate this feedback locally, and adjust existing prediction models accordingly to fit their population and setting (Spengler et al., 1995). Finally, these prediction models should be converted to easy-to-use computerized clinical decision support tools that can be used to fuel the dialogue between the clinician and the patient (Katsikopoulos et al., 2008; Roshanov et al., 2013). In this way, it turns statistical prediction into an empowering platform for ongoing shared decision making.

Funding

This research was funded by the research institute of Experimental Psychopathology (EPP), the Netherlands, and the Academic Community Mental Health Centre (RIAGG, now METggz Maastricht) in Maastricht, the Netherlands. Both institutes had no involvement in the collection, analysis and interpretation of the data or in the writing of this article and decision to submit this article for publication.

Contributors

MH and FP designed the study. LL, MH and FP were responsible for the collection and management of study data. SB, LL, MH and FP

drafted the analysis plan, SB did the statistical analysis. SB wrote the first draft of the manuscript. All authors were involved in data interpretation, critically revising the manuscript and approved the final version.

Declaration of Competing Interest

All authors declare that they have no competing interests.

Acknowledgements

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

References

Ægisdóttir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Anderson, L.A., Cook, R.S., Nichols, C.N., Lampropoulos, G.K., Walker, B.S., Cohen, G., 2006. The meta-analysis of clinical judgment project: fifty-six years of accumulated research on clinical versus statistical prediction. Counsel. Psychol. 34, 341–382.
Austin, P.C., Tu, J.V., 2004. Bootstrap methods for developing predictive models. Am. Stat. 58, 131–137.
Barber, J.P., Muenz, L.R., 1996. The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: empirical findings from the treatment for depression collaborative research program. J. Consult. Clin. Psychol. 64, 951–958.
Beck, A.T., Rush, A.J., Shaw, B.F., Emery, G., 1979. Cognitive Therapy of Depression. Guilford Press, New York.
Beck, A.T., Steer, R., Brown, G.K., 1996. Beck Depression Inventory II: Manual. Harcourt Brace, Boston.
Beijers, L., Wardenaar, K.J., van Loo, H.M., Schoevers, R.A., 2019. Data-driven biological subtypes of depression: systematic review of biological approaches to depression subtyping. Mol. Psychiatry 24, 888–900.
Bell, I., Mellor, D., 2009. Clinical judgements: research and practice. Austr. Psychol. 44, 112–121.
Carter, J.D., Luty, S.E., McKenzie, J.M., Mulder, R.T., Frampton, C.M., Joyce, P.R., 2011. Patient predictors of response to cognitive behaviour therapy and interpersonal psychotherapy in a randomised clinical trial for depression. J. Affect. Disord. 128, 252–261.
Cohen, Z.D., DeRubeis, R.J., 2018. Treatment selection in depression. Ann. Rev. Clin. Psychol. 14, 209–236.
Cooper, A.A., Conklin, L.R., 2015. Dropout from individual psychotherapy for major depression: a meta-analysis of randomized clinical trials. Clin. Psychol. Rev. 40, 57–65.
Cuijpers, P., Andersson, G., Donker, T., van Straten, A., 2011. Psychological treatment of depression: results of a series of meta-analyses. Nord. J. Psychiatry 65, 354–364.
Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C.H., Cipriani, A., Furukawa, T.A., 2020a. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry 19, 92–107.
Cuijpers, P., Reijnders, M., Huibers, M.J., 2019. The role of common factors in psychotherapy outcomes. Ann. Rev. Clin. Psychol. 15, 207–231.
Cuijpers, P., Stringaris, A., Wolpert, M., 2020b. Treatment outcomes for depression: challenges and opportunities. Lancet Psychiatry.
Dawes, R.M., 2005. The ethical implications of Paul Meehl's work on comparing clinical versus actuarial prediction methods. J. Clin. Psychol. 61, 1245–1255.
Delgadillo, J., Gellatly, J., Stephenson-Bellwood, S., 2015. Decision making in stepped care: how do therapists decide whether to prolong treatment or not? Behav. Cogn. Psychother. 43, 328–341.
Delgadillo, J., Lutz, W., 2020. A development pathway towards precision mental health care. JAMA Psychiatry.
DeRubeis, R.J., 2019. The history, current status, and possible future of precision mental health. Behav. Res. Ther. 123, 103506.
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, e83875.
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. Br. J. Psychiatry 192, 124–129.
Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. J. Socio-Econ. 40, 35–42.
Garb, H.N., 2000. Computers will become increasingly important for psychological assessment: not that there's anything wrong with that. Psychol. Assess. 12, 31.
Garb, H.N., 2005. Clinical judgment and decision making. Ann. Rev. Clin. Psychol. 1, 67–89 2005.
Garge, N.R., Bobashev, G., Eggleston, B., 2013. Random forest methodology for model-based recursive partitioning: the mobforest package for R. BMC Bioinf. 14, 1–8.
Grove, W.M., Meehl, P.E., 1996. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. Psychol. Public Policylaw 2, 293.
Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., Nelson, C., 2000. Clinical versus mechanical prediction: a meta-analysis. Psychol. Assess. 12, 19.
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, e0140771.
Joyce, P.R., McKenzie, J.M., Carter, J.D., Rae, A.M., Luty, S.E., Frampton, C.M., Mulder, R.T., 2007. Temperament, character and personality disorders as predictors of response to interpersonal psychotherapy and cognitive-behavioural therapy for depression. Br. J. Psychiatry 190, 503–508.
Katsikopoulos, K.V., Pachur, T., Machery, E., Wallin, A., 2008. From meehl to fast and frugal heuristics (and back) new insights into how to bridge the clinical—actuarial divide. Theo. Psychol. 18, 443–464.
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. Depress. Anxiety 35, 330–338.
Klerman, G.L., Weissman, M.M., Rounsaville, B.J., Chevron, E.S., 1984. Interpersonal Psychotherapy for Depression. Basis Books, New York.
Knaup, C., Koesters, M., Schoefer, D., Becker, T., Puschner, B., 2009. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. Br. J. Psychiatry 195, 15–22.
Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D.B., Paolini, M., Chisholm, K., Kambeitz, J., Haidl, T., 2018. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA Psychiatry 75, 1156–1172.
Kroenke, K., 2007. Efficacy of treatment for somatoform disorders: a review of randomized controlled trials. Psychosom. Med. 69, 881–888.
Lemmens, L., Arntz, A., Peeters, F., Hollon, S., Roefs, A., Huibers, M., 2015. Clinical effectiveness of cognitive therapy v. interpersonal psychotherapy for depression: results of a randomized controlled trial. Psychol. Med. 45, 2095–2110.
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.
Lorenzo-Luaces, L., DeRubeis, R.J., Bennett, I.M., 2015. Primary care physicians' selection of low-intensity treatments for patients with depression. Fam Med 47, 511–516.
Markowitz, J.C., Skodol, A.E., Petkova, E., Cheng, J., Sanislow, C.A., Grilo, C.M., Gunderson, J.G., McGlashan, T.H., 2007. Longitudinal effects of personality disorders on psychosocial functioning of patients with major depressive disorder. Journal of Clinical Psychiatry 68, 186.
Meehl, P.E., 1954. Clinical versus statistical prediction: a theoretical analysis and a review of the evidence.
Meehl, P.E., 1986. Causes and effects of my disturbing little book. J. Personal. Assess. 50, 370–375.
Nickerson, R.S., 1998. Confirmation bias: a ubiquitous phenomenon in many guises. Rev. Gen. Psychol. 2, 175–220.
Papakostas, G.I., Fava, M., 2010. Pharmacotherapy for Depression and Treatment-Resistant Depression. World Scientific.
Perna, G., Grassi, M., Caldirola, D., Nemeroff, C., 2018. The revolution of personalized psychiatry: will technology make it happen sooner? Psychol. Med. 48, 705–713.
Picard, R.R., Cook, R.D., 1984. Cross-validation of regression models. J. Am. Stat. Asso. 79, 575–583.
Richards, P.M., Geiger, J.A., Tussey, C.M., 2015. The dirty dozen: 12 sources of bias in forensic neuropsychology with ways to mitigate. Psychol. Injury law 8, 265–280.
Rizopoulos, D., Rizopoulos, M.D., 2009. Package 'boot step AIC'.
Roshanov, P.S., Fernandes, N., Wilczynski, J.M., Hemens, B.J., You, J.J., Handler, S.M., Nieuwlaat, R., Souza, N.M., Beyene, J., Van Spall, H.G., 2013. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. Bmj 346, f657.
Rush, A.J., Trivedi, M.H., Wisniewski, S.R., Nierenberg, A.A., Stewart, J.W., Warden, D., Niederehe, G., Thase, M.E., Lavori, P.W., Lebowitz, B.D., 2006. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report. Am. J. Psychiatry 163, 1905–1917.
Ryder, A.G., Quilty, L.C., Vachon, D.D., Bagby, R.M., 2010. Depressive personality and treatment outcome in major depressive disorder. J. Pers. Disord. 24, 392–404.
Simon, G.E., Perlis, R.H., 2010. Personalized medicine for depression: can we match patients with treatments? Am. J. Psychiatry 167, 1445–1455.
Spengler, P.M., Strohmer, D.C., Dixon, D.N., Shivy, V.A., 1995. A scientist-practitioner model of psychological assessment: implications for training, practice and research. Counsel. Psychol. 23, 506–534.
Stekhoven, D.J., Bühlmann, P., 2012. Miss forest—non-parametric missing value imputation for mixed-type data. Bioinformatics 28, 112–118.
Swift, J.K., Greenberg, R.P., Tompkins, K.A., Parkin, S.R., 2017. Treatment refusal and premature termination in psychotherapy, pharmacotherapy, and their combination: a meta-analysis of head-to-head comparisons. Psychotherapy 54, 47.
van Bronswijk, S.C., DeRubeis, R.J., Lemmens, L.H., Peeters, F.P., Keefe, J.R., Cohen, Z.D., Huibers, M.J., 2019. Precision medicine for long-term depression outcomes using the personalized advantage index approach: cognitive therapy or interpersonal psychotherapy? Psychol. Med. 1–11.
Warden, D., Trivedi, M.H., Carmody, T., Toups, M., Zisook, S., Lesser, I., Myers, A., Kurian, K.R.B., Morris, D., Rush, A.J., 2014. Adherence to antidepressant combinations and monotherapy for major depressive disorder: a CO-MED report of measurement-based care. J. Psychiatr. Pract.® 20, 118–132.
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. J. Clin. Psychiatry 77, e1584–e1590.