Personalized Medicine and Cognitive Behavioral Therapies for Depression: Small Effects, Big Problems, and Bigger Data

Lorenzo Lorenzo-Luaces[1] · **Allison Peipert**[1] · **Robinson De Jesús Romero**[1] · **Lauren A. Rutter**[1] · **Natalie Rodriguez-Quintana**[2]

Abstract

Cognitive-behavioral therapies (CBTs) are the most widely studied form of psychotherapy for disorders like depression and anxiety. Nonetheless, there is heterogeneity in response to CBTs vs. other treatments. Researchers have become increasingly interested in using pre-treatment individual differences (i.e., moderators) to match patients to the most effective treatments for them. Several methods to combine multiple variables to create precision treatment rules (PTRs) that identify subgroups have been proposed. We review the rationale behind multivariable PTRs as well as the findings of studies that have used different PTRs. We identify conceptual and methodological issues in the literature. Multivariable treatment assignment is a promising avenue of research. Nonetheless, effect sizes appear to be small and most of the samples that have been used to study these questions have been grossly underpowered to detect small effects. We recommend researchers explore multivariable treatment selection strategies, particularly those resembling risk stratification, in heterogeneous samples of patients undergoing low-intensity CBTs vs. realistic minimal controls.

Keywords Personalized medicine · Machine learning · Cognitive-behavioral therapy · Depression

So-called mental disorders account for a substantial proportion of the disability attributable to health conditions (Whiteford et al. 2013). Depressive disorders account for a substantial amount of that disability, in part due to how common and disabling they are (Patel et al. 2016). An often-cited feature of depression is its heterogeneous symptom

✉ Lorenzo Lorenzo-Luaces
  lolorenz@iu.edu

1 Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

2 Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

presentation (Fried and Nesse 2015; Zimmerman et al. 2015). Depression is also heterogeneous in its prognosis (Lorenzo-Luaces 2015; Monroe and Harkness 2005, 2011). While most cases in naturalistic samples may remit within a 3–6-month period, a minority of individuals have a chronic or unremitting course. Roughly half of cases that recover do not go on to relapse, but, of individuals who relapse, repeated episodes are common (Lorenzo-Luaces 2015; Monroe and Harkness 2005, 2011).

Cognitive-behavioral therapies (CBTs) have been the most widely studied psychological interventions for depression as well as for other common mental disorders (Barth et al. 2016; Lorenzo-Luaces et al. 2020b). We refer to CBTs as a family of interventions which may include mostly components aimed at changing cognitions (e.g., cognitive-processing therapy (CPT)), mostly components aimed at changing behaviors (e.g., behavioral activation (BA)), or treatments that combine these components (DeRubeis and Lorenzo-Luaces 2017). CBTs appear equally effective to antidepressant medications in the acute phase of treatment (Weitz et al. 2015) as well as to other psychotherapies (Barth et al. 2016). Over the long term, CBTs appear to have a relapse-prevention component that may be as effective as keeping patients on medications and superior to withdrawing medications after acute treatment (Cuijpers et al. 2013). Long-term comparisons between CBTs and other treatments are fewer and far between. However, the existing data suggests that if there are long-term outcome differences between CBTs and other therapies like interpersonal therapy (IPT) or psychodynamic therapy (PDT), the differences are likely small and of questionable clinical significance (Driessen et al. 2013; Lemmens et al. 2019).

Interest in Treatment Outcome Moderators

Given that no treatment appears superior to another and that none of the effective treatments for depression are effective for all patients, researchers have hypothesized that there are individual differences in response to treatments (Cohen and DeRubeis 2018). This hypothesis is difficult to explore given the inability to treat the same patient with two different treatments without introducing serious confounds (e.g., the effect of ordering of treatments). Researchers and clinicians assume that there are stable individual variables that can be assessed pre-treatment and that may predict differential treatment outcomes between CBTs and other treatments.

The study of patient variables that predict treatment outcomes generally concerns one of two sets of findings: finding of prognostic variables, also known as predictors, as well as of prescriptive variables, sometimes known as moderators (Fournier et al. 2009). Predictor or prognostic variables predict overall treatment outcomes or treatment outcomes irrespective of treatment (e.g., in multiwave studies). Moderator or prescriptive variables predict *differential* outcomes in comparisons of two or more treatments. They are usually detected by the statistical interaction of treatment and a baseline variable (Kraemer et al. 2002). The moderator effects may be disordinal, sometimes also known as "cross-over," if a treatment contrast effect is reversed in one value of a moderator variable vs. another. For example, if age were known to be a disordinal moderator of the difference between CBT and antidepressants, it may indicate that while the treatments appear equally effective on average CBT is superior to antidepressants for young children whereas antidepressants are superior to CBTs for older

adults. In an ordinal moderator effect, the treatment differences never fully reverse. For example, if age was an ordinal moderator of the difference between CBT and antidepressants, it may indicate that there are no differences between CBT and antidepressants for young children while antidepressants are superior to CBTs for older adults.

At least four groups of variables have been thoroughly explored as pre-treatment moderators of response to CBT vs. other treatments: symptoms (e.g., psychomotor retardation), psychological vulnerabilities relevant to treatment mechanisms (e.g., high levels of cognitive distortions), preferences for the treatment (e.g., preferring CBTs to antidepressants), and demographics (e.g., age).

DSM Symptoms

The DSM differential diagnosis of depression (APA 2013) includes the distinction between a recurrent course vs. a first episode, the diagnosis of persistent depressive disorder if an episode lasts two or more years, and nine specifiers. As well, diagnosing depression involves considering entirely different diagnostic classes for depression occurring in the context of mania or hypomania (i.e., bipolar disorders), independent psychotic episodes (i.e., schizoaffective disorder), or medications and general medication conditions. While the field has antipathy towards the use of the DSM to parse heterogeneity (Kotov et al. 2017; Olbert et al. 2014), some of its attempts to reduce heterogeneity could be considered successful. In general, psychotic depression (Perez et al. 2018), depression with catatonia (Ungvari et al. 2001), and persistent depressive disorders (Cuijpers et al. 2010; Hollon et al. 2014; Lorenzo-Luaces et al. 2020c) appear to have poorer overall outcomes than their counterparts (i.e., non-psychotic, non-catatonic, and non-chronic depression). Their presence in a patient may suggest the need for pharmacological interventions either in addition to or instead of CBTs (Hollon et al. 2014; Lorenzo-Luaces et al. 2020c). While recurrent depression may not be a predictor of acute outcomes (Fournier et al. 2009), it tends to predict a more recurrent course and may thus be useful for treatment assignment when considering long-term outcomes in depression (Bockting et al. 2015). A commonality of the disorder features that predict outcomes is that they are specifiers indicating the history or course of depression (e.g., chronic or recurrent) (Lorenzo-Luaces 2015) or that they add symptoms that appear particularly severe and which may predict a poorer prognosis (e.g., catatonia, psychosis, (hypo)mania).

A priori subtypes which may be more qualitative in describing different symptom patterns like melancholic and atypical depression do not appear to reliably predict treatment outcomes (Cuijpers et al. 2017). Moreover, these subtypes may introduce more heterogeneity into symptom presentations, possibly because they consider symptoms that go beyond the major depressive episode (MDE) criteria (Fried et al. 2020; Lorenzo-Luaces et al. 2020a). The mixed features and anxious-distress specifiers recently added to the DSM appear to predict worse outcomes in treatment relative to their counterparts (Gaspersz et al. 2017; Pae et al. 2012). However, it is unclear to what extent these specifiers can be used for treatment matching, especially the anxious-distress specifier which 70–75% of individuals with MDD meet criteria for (Hasin et al. 2018). Similarly, it is unclear to what degree subtypes that

characterize the patterns (e.g., seasonal-affective or peripartum) validly predict differential outcomes (Baumeister and Parker 2012).

Researchers have also used atheoretical approaches to identify constellations of DSM symptoms that may predict differential response to interventions. For example, in an individual patient data (IPD; $n = 1070$) meta-analysis, Boschloo et al. (2019) compared the magnitude of symptom change for the individual items of the Hamilton Rating Scale for Depression (HRSD) for patients randomized to CBT vs. antidepressants. Five of the 17 HRSD symptoms—depressed mood, feelings of guilt, suicidal thoughts, psychic anxiety (i.e., worry), and general somatic symptoms—changed more in antidepressants than CBT. Although the individual effect sizes were small (d s = 0.13–0.16), individuals who scored highly (75th percentile) in these symptoms experienced better outcomes in antidepressants than CBT ($d = 0.30$). Thus, while overall depression severity may not moderate differences between CBT and antidepressants (Weitz et al. 2015), there is some evidence that considering a specific severity profile may be useful in allocating patients to CBTs vs. antidepressants.

It should be observed that in the study by Boschloo et al. (2019), antidepressants had a small degree of superiority to CBT. The finding was not that symptoms allow for qualitative matching of patients to treatments where some patients experience better outcomes in CBT vs. antidepressants while others experience better outcomes in antidepressants than in CBT. Instead, the authors were able to find the specific severity profile where antidepressants may outperform CBT. A more recent analysis of similar data (Kappelmann et al. 2020), however, failed to find that there is any utility to differential treatment assignment based on baseline depression symptoms.

The question of whether individual DSM symptoms have a value in treatment assignment has been a widely studied question in depression research (Kessler et al. 2017), probably due to how often these data are available in clinical trials. It appears that symptoms that go beyond the depression criteria like psychosis and hypomania may be predictors of treatment response. But there is no consistent or strong evidence that the symptoms of depression, at least the ones that are measured in the symptom scales most often administered in randomized controlled trials (RCTs), are moderators of response to CBT vs. other treatments.

Compensation or Capitalization?

Intuitively, one may hypothesize that individuals benefit more from a treatment if they have the psychological vulnerability the treatment targets. For example, one may expect that a patient with reasonably healthy interpersonal relations but relatively high levels of distorted cognitions may benefit more from CBT than from other treatments like IPT and PDT. The tacit understanding of this hypothesis is that such an individual would have depression that is maintained by maladaptive cognitions and behaviors, and "since" CBTs target cognitive and behavioral vulnerabilities (Lorenzo-Luaces et al. 2015; Lorenzo-Luaces et al. 2016), the treatments should be superior to alternatives that focus on other vulnerabilities. This "compensation" approach to therapy has been contrasted to a "capitalization" approach wherein one considers a patient well-suited to a treatment that fosters skills they already have. In the preceding

example with a patient who has good interpersonal relationships, IPT may be a suitable alternative. PDT may be best for a patient with relatively good insight over unconscious dynamics and so forth.

At least one randomized controlled trial (RCT) by Jennifer Cheavens has compared the compensation approach to the capitalization approach in depression, finding that the capitalization approach produced faster results (Cheavens et al. 2012). In a trial exploring the implementation of CBTs for generalized anxiety disorder, Cristoph Flückiger and colleagues also found results that support the superiority of the capitalization over the compensation model (Flückiger et al. 2016). Moderator analyses in studies comparing CBT to other treatments also support a capitalization model. For example, re-analyzing data from an RCT, Huibers et al. (2015) found that IPT was superior to CBT for patients with cognitive problems, whereas CBT tended to be superior to IPT among patients with an interpersonal "self-sacrificing" style, more paranoid symptoms, and more life events in the past year. Other moderator findings in psychotherapy research support the "capitalization" approach. For example, in a small RCT ($N = 55$), McBride et al. (2006) found that patients with more attachment avoidance experience better outcomes in CBT than in IPT. A similar pattern of findings was reported by Joyce et al. (2007) who found that avoidant and schizoid symptoms predicted poorer response to IPT ($n = 87$) but not CBT ($n = 80$), suggesting that patients who report higher levels of interpersonal problems experience better outcomes in CBT than IPT.

Preference for Treatment

Common sense would suggest that if an individual has a preference for a psychological treatment, they may be more likely to complete it and benefit from it, relative to a treatment they do not prefer. A recent meta-analytic review (Windle et al. 2020) of 29 randomized and non-randomized studies involving 5294 patients provided evidence to suggest that receiving a preferred psychosocial treatment was associated with lower likelihood of dropping out of treatment (risk ratio (RR) = 0.62, 95% confidence interval (CI) = 0.48–0.80). However, receiving one's preferred treatment was not associated with improvement in depression or anxiety ($d = 0.01$, 95 CI = − 0.18–0.20). Given that treatment preferences have also been widely studied variables, this finding should temper interest in preference as a variable that can be used for differential treatment allocation. Nonetheless, in specific contexts where dropout is high, for example with guided CBT self-help (Cuijpers et al. 2019), knowing treatment preferences may prove important.

Demographics

Most treatment trials assess demographics, which are sometimes mandatory variables to report (Polo et al. 2019). Accordingly, demographics have also been studied extensively as moderators as well as predictors of treatment outcomes.

Gender

One of the most comprehensive accounts of the extent to which gender moderates response to CBTs vs. other treatments was published by Cuijpers et al. (2014). They conducted an individual patient data (IPD) meta-analysis of 14 studies of CBT for depression vs. antidepressants ($N = 1766$) and tested whether treatment outcomes varied according to reported sex overall (i.e., if sex has prognostic significance) or by treatment (i.e., if sex is a moderator of outcomes in CBT vs. antidepressants). There was no evidence that overall outcomes varied between men and women nor that there were differences by gender between CBTs and antidepressants, CBTs and placebos, or antidepressants and placebos. In another IPD meta-analysis of guided Internet-based therapies ($N = 4889$), most of which were CBTs, vs. controls (e.g., waiting lists, care as usual), the authors also failed to find that sex moderated treatment outcomes (Karyotaki et al. 2018). This is not to say that gender has no predictive value in relation to treatment outcomes. For example, in a small study exploring moderators of the effects of adherence to different aspects of CBT, gender moderated the effects between cognitive methods and subsequent symptom change (Sasso et al. 2015). Specifically, for men, cognitive methods did not predict symptom change while, for women, cognitive methods predicted decreased symptoms change. Thus, it is possible that sex predicts response to specific CBT techniques per se while not to whether the treatment will be superior to a control condition or antidepressants.

Age

In a recent meta-analysis summarizing data from 366 trials, Cuijpers and colleagues reported that "[psychotherapy] effect sizes are small in childhood, become larger in adolescents, and grow further in early adulthood, before decreasing to more modest effect sizes in the rest of the adult population" (Cuijpers et al. 2020). Although the authors do not report CBT-specific data, a substantial proportion of the studies analyzed were CBT studies, and the age effects did not appear to vary across treatment modalities. In the IPD meta-analysis of guided self-help, mostly CBTs, by Karyotaki et al. (2018), age also moderated treatment outcomes in comparisons relative to waiting lists, care as usual, or other controls (e.g., psychoeducation). Older adults were more likely to experience symptom remission or response in guided Internet-based therapies than in control conditions. When considering these findings, it is important to consider that the response to non-specific controls (e.g., placebos) appears particularly large in youth with depression (D. Cohen et al. 2008; Locher et al. 2017), which may make it difficult to find treatment-specific effects in treatment studies with youth. It is unclear, however, whether age moderates outcomes when contrasting CBTs to other treatments like antidepressants or other psychotherapies.

Mutivariable Algorithms

Taken together, the existing data suggest that many of our a priori assumptions about who should respond best to CBT are wrong: variables that are easy to assess like sex do not moderate outcomes, depressive symptoms and specifiers do not appear to be

reliable moderators, preferences do not predict symptom change, and it is likely that patients perform best in CBTs when they already have some of the skills the therapy is meant to foster. While the presence of psychosis, manic symptoms, and a chronic/ recurrent course, along with a more severe profile, may suggest than an additional treatment needs to be considered, there is little else in the way of using patients' information to decide treatment allocation.

Sidestepping theory, researchers have used more exploratory approaches to identify variables that may moderate treatment outcomes in CBT vs. other treatments by mining existing databases to identify moderators of outcomes (Cohen and DeRubeis 2018). These efforts fall under the umbrella of "personalized," also sometimes known as "precision" medicine.

While some moderators have been reported to predict outcomes in the literature more than once (Kessler et al. 2017), the field has yet to converge on a single moderator variable that may be useful for treatment allocation. DeRubeis (2019) conjectured that there probably is no such single variable, given that it would need to be a variable that accounts for a very large effect by itself, while also being something that has escaped the notice of the field.

In the past 20 years, and especially in the last 7, several authors have proposed combining variables that may predict differential outcomes to mental health interventions (Barber and Muenz 1996; Beutler et al. 1991; Cloitre et al. 2016; DeRubeis et al. 2014a; Kraemer 2013; Lorenzo-Luaces et al. 2017; Lutz et al. 2006; Petkova et al. 2020; Wallace et al. 2013). The resulting models have taken various names including composite/combined moderator (M*), the Personalized Advantage Index (PAI), precision treatment rules (PTRs), generated effect modifiers (GEMs), and aptitude-by-treatment interactions (ATIs). These methods for estimating the effect of treatments for potential subgroups have been so popular that Hollon et al. (2019) described how they could "revolutionize the field" (p. 260) when applied to large datasets. Following Kessler (2018), we refer to these as precision treatment rules (PTRs), which is a term that does not imply following a specific method (e.g., the M* approach suggested by Kraemer (2013)).

Personalized Advantage Index

In one application of a multivariable PTR, DeRubeis et al. (2014a) used data from a RCT that compared antidepressants ($n = 120$) to CBT ($n = 60$) and found that both were superior to placebo ($n = 60$) but neither active treatment was superior to the other (DeRubeis et al. 2005). Across several studies, the authors had identified predictive variables, predicting overall response, as well prescriptive variables, predicting response to each treatment (Fournier et al. 2008; Fournier et al. 2009; Leykin et al. 2007). Being unemployed, being married, not having a personality disorder, having tried antidepressants previously, and more negative life events predicted better response to CBT than to antidepressants. By contrast, having a personality disorder predicted better response to antidepressants than to CBT. Using the data from these prior publications, we (DeRubeis et al. 2005) combined all the variables in a regression model and generated, for each patient, a predicted score for CBT and a predicted score for antidepressants. Then, we calculated the difference between predicted outcome in CBT vs. antidepressants for each patient and dubbed it the "Personalized Advantage

Index" (PAI). For example, a patient with a predicted antidepressant score of 20 on the HRSD and of 10 on CBT is predicted to perform better on CBT than antidepressant. The signed difference between the two estimates indicate which treatment is most effective and by how much. The absolute score provides an index on the size of the effect of being matched to one's preferred treatment, irrespective of what the treatment is. Because the study was an RCT, we could also identify outcomes for patients who, randomly, had been assigned to their model-optimal treatment vs. those who were assigned to a treatment that was not optimal to them. Overall, patients assigned to their optimal treatment had superior outcomes than patients assigned to their non-optimal treatment ($d = 0.28$, $p = 0.09$). By conventional statistical standards (e.g., p value < 0.05), the difference in outcome was small. Nonetheless, the differences between CBT and antidepressant that had been reported in the literature had been minimal, suggesting that the PAI approach is a promising way to improve overall outcomes. When the analyses were limited to patients who had the highest PAIs, the effect of assigning to the "optimal" treatment was more pronounced than in the full sample.

Is It the Size or Is It the Fit?

The study by DeRubeis et al. (2014b) is one of several efforts to combine moderator variables in RCT so as to quantify the effects of PTRs. We reviewed the literature that has been produced by studies attempting to use multivariable prediction efforts. We did so by cross-referencing the articles that had been the first to suggest a method of combining moderator variables to predict outcomes (Barber and Muenz 1996; Beutler et al. 1991; Cloitre et al. 2016; DeRubeis et al. 2014a; Kraemer 2013; Lorenzo-Luaces et al. 2017; Lutz et al. 2006; Wallace et al. 2013) on Scopus. Two raters (AP and RDJR), consulting the first author, identified studies if they used (a) a multivariable prediction model, (b) exploring response to CBT, and (c) predicted depression outcomes. Our aim was *not* to conduct a systematic review of the literature and the results of the search may be better-interpreted as a scoping review. Other authors have used systematic methods to review the literature on prediction on treatment response (Aafjes-van Doorn et al. 2020; Kessler et al. 2017; Lee et al. 2018; Rekkas et al. 2020), and several others have also published conceptual reviews on models in mental health in general or even depression specifically (Cohen and DeRubeis 2018). Instead, our aim was to conduct a general literature review to identify the pattern of results that have been reported to date.

Reviewing the literature on the effects of using PTRs (see Table 1) revealed several patterns. First, there was heterogeneity in how results were presented. For each study, we initially wanted to identify the overall effects of multivariable treatment assignment, which is usually assigned post hoc, as well as whether specific advantages were reported for one treatment than for the other (e.g., whether the PTR was a better predictor of outcomes in CBT vs. the other treatment), the variables that informed the prediction model, and what percentage of the sample would have been predicted to have treatment recommendations different than either randomization or assignment to the more effective treatments. We were unable to find this information for most of the studies so we limited our presentation of the findings to include the treatments under

Table 1 Studies exploring multivariable prediction of response to CBT vs. other treatments

Study	Intervention	PTR method	Number	Effect size
Barber and Muenz (1996)	CBT vs. IPT	Matching factor	CBT: $n=37$, IPT: $n=47$	CBT: $d=0.89$ IPT: $d=0.98$
Cohen et al. (2020)	CBT vs. PDT	PAI-Multi	CBT: $n=75$, PDT: $n=92$	$d=0.21$
Delgadillo et al. (2017)	LIT, HIT, LIT + HIT	PI	LIT-Test $=269$, HIT-Test $=131$, LIT + HIT $=140$	HIT vs. LIT: OR $=2.33$
Delgadillo and Gonzalez Salas Duhne (2020)	CBT vs. CfD	PAI-EN	CBT-Train: $n=929$, CfD-Train: $n=156$, CBT-Test: $n=175$, CfD-Test: $n=175$	$g=0.26$
Eskildsen et al. (2020)	CBT vs. tCBT	PAI-RF/BS	CBT: $n=143$ tCBT: $n=148$	No moderators found
Friedl et al. (2020)	CBT vs. CBT-EE	PAI-BMA	CBT: $n=72$, CBTEE: $n=77$	$d=0.16$
Huibers et al. (2015)[a]	CBT vs. IPT	PAI-OLS	CBT: $n=69$, IPT $=65$	$d=0.51$, CT: $d=0.66$, IPT: $d=0.34$
van Bronswijk et al. (2020)[a]—StepD	CBT vs. IPT	PAI-RF/BS	CBT: $n=76$, IPT: $n=75$	Train: $d=0.57$, test: $d=0.16$
van Bronswijk et al. (2020)*—FreqMech	CBT vs. IPT	PAI-RF/BS	CBT $=98$, IPT $=102$	Train: $d=0.20$, test: $d=0.27$
van Bronswijk et al. (2019) (follow-up)	CBT vs. IPT	PAI-RF/BS	CBT: $n=76$, IPT: $n=75$	$d=0.23$, CT: $d=0.42$, IPT: $d=-0.11$
Lopez-Gomez et al. (2019)	CBT vs. PPI	PAI-RF/EN	CBT: $n=66$, PPI: $n=62$	$d=0.24$, CBT: $d=0.21$, PPI: $d=0.37$
Lorenzo-Luaces et al. (2017)	CBT vs. TAU CBT vs. BT	PI	CBT: $n=208$, BT: $n=179$, TAU: $n=234$	CBT vs. TAU: OR $=1.90$ CBT vs. BT: OR $=1.88$
Lorenzo-Luaces et al. (2020d)	CBT vs. COMB CBT vs. ADM	PI	Placebo/train: $n=112$, CBT/test: $n=111$, ADM/test: $n=109$, COMB/test: $n=107$	COMB vs. CBT: $d=0.68$, ADM vs. CBT: $d=0.44$
Lutz et al. (2006)	CBT vs. IT	NN	CBT: $n=262$, IT: $n=356$	NR
	PST vs. CAU	M*-LASSO	PST: $n=89$, CAU: $n=84$	$r=0.72$

Table 1 (continued)

Study	Intervention	PTR method	Number	Effect size
Otero et al. (2015)				
Schwartz et al. (2020)	CBT vs. PDT	PAI-RF/BS	CBT-Train: $n=654$, PDT-Train: $n=312$, CBT-Test: $n=281$, PDT-Test: $n=132$	Train: $d=0.14$, test: $d=0.09$
Serbanescu et al. (2020)	CBASP vs. SP	M*	CBASP: $n=149$, SP: $n=131$	$r=0.34$, CBASP: $d=-0.57$, SP: $d=0.29$
Wallace et al. (2013)	IPT vs. ADM	M*-OLS	IPT: $n=149$, ADM: $n=142$	$d=0.31$, IPT: $d=0.35$, ADM: $d=-0.39$

Effect sizes correspond to a standardized effect given to summarize the effects of using the treatment allocation rule or observed advantage among those randomized to the indicated treatment

a analyses used the same training dataset, * analyses used a different dataset than Huibers et al. (2015), CBT cognitive behavioral therapy, IPT interpersonal therapy, CBT-EE CBT with exposure and emotion-focused elements, PDT psychodynamic therapy, LIT low-intensity treatment, HIT high-intensity treatment, PAI personalized advantage index, CfD person-centered counseling, ADM antidepressant medication, NR not reported, PPI positive psychological intervention, PST problem-solving therapy (PST), PI prognostic index, TAU treatment as usual, BT brief therapy, M* composite moderator, OR odds ratio, COMB combination of CBT and ADM, NN nearest neighbors, CAU care as usual, BA behavioral activation, CBASP cognitive behavioral system of psychotherapy, Multi pooling together multiple predictions, OLS ordinary least squares, RF random forests, BS bootstrapped regressions, EN elastic net

consideration, the method of creating a PTR, sample sizes per study arm, and a measure of effect size reported for the study.

According to researchers, one of the promises of multivariable treatment assignment is that combining information from multiple moderators can lead to larger treatment differences than only using a single moderator (Cohen and DeRubeis 2018; DeRubeis et al. 2014a; Kraemer 2013; Wallace et al. 2013). While this claim appears to be true prima facie, it nonetheless needs to be evaluated empirically. Surveying the studies that report on multivariable treatment assignment, it appears that the effect sizes reported have been highly variable in magnitude. Some studies report very large effect sizes. For example, Otero et al. (2015) explored moderators of response to a problem-solving therapy (PST) vs. usual care for preventing depression among caregivers. The authors found that a composite moderator as per the method suggested by Kraemer (2013) interacted with treatment conditions to predict treatments outcomes, with a large effect ($r = 0.72$, 95% CI 0.43–1.01). Others report much smaller effect sizes (Z. D. Cohen et al. 2020; Friedl et al. 2020; van Bronswijk et al. 2019). For example, in an exploration of moderators of response to CBT vs. PDT for depression, Schwartz et al. (2020) reported on a multivariable PTR. In the subsection of data in which the PTR was developed, patients who had been randomly assigned to the model-predicted optimal treatment experienced better outcomes than who had been assigned the model-predicted suboptimal treatment, though this difference was small ($d = 0.14$, 95% CI 0.02–0.27). In another study of 291 patients assigned to group CBT or transdiagnostic group CBT, Eskildsen et al. (2020) could not identify moderators of response and thus could not create a PTR.

It is not sensible to outright dismiss an effect size by using an arbitrary cutoff (e.g., $d \geq 0.20$). Nonetheless, it is important to consider the magnitude of the effect sizes that have been reported. The overall effect sizes that have been reported may be indicative of the upper limits of the benefit obtained from using multivariable PTRs. This is because there are reasons to believe that these small effect sizes are inflated. Secondary analyses of an RCT by Lemmens et al. (2015), which compared IPT and CBT, illustrate this point. Huibers et al. (2015) constructed a multivariable PTR from data from that RCT ($N = 134$). Examining outcomes in individuals who had been assigned to their model-predicted optimal treatment vs. their model-predicted non-optimal treatment suggested that there were medium differences that may be explained by treatment matching ($d = 0.51$). In a subsequent publication (Bruijniks et al. 2020), we used machine-learning methods to conduct a similar analysis, still finding large effects of the PTR on those assigned to their optimal vs. non-optimal treatment ($d = 0.57$). Nonetheless, when we applied that prediction model to data from another trial (van Bronswijk et al. 2020), the observed advantage of the multivariable moderator was small ($d = 0.16$); the effect size shrank over 77% of its original size.

Statistical Shrinkage

The concept of statistical shrinkage refers to the idea, in lay man's terms, that models tend to have a better fit in the data in which they are developed than in new data. It is closely tied to concerns about overfitting, including fitting spurious relations. This type of overfitting may be more likely to occur when one uses a dataset to identify variables

to build a model (e.g., if they pass a $p < 0.05$ threshold) and then tests the model in the same data, which is sometimes known as "double-dipping" (Kriegeskorte et al. 2009).

In a prior publication, we assumed that treatment selection may be more important for patients who are predicted to have greater predicted differences between treatments (DeRubeis et al. 2014b). Similarly, DeRubeis (2019) noted that while many efforts to identify patient subgroups are limited by small sample sizes, small subsets of patients may be predicted to experience large benefits from treatment selection. While these recommendations are sensible, if the methods researchers use to compute the effects of multivariable PTRs are prone to overfitting, then the so-called large effects reported are not known to be large; they only appear to be so because of overfitting. One technique that has been developed to decrease the likelihood of overfitting is regularization. Regularization shrinks regression coefficients so they are more "conservative," sometimes shrinking regression coefficient that are small to zero. The concept of regularization is at times counter-intuitive because scientists are trained to explore large effects and clinicians as well are interested in large effects. Nonetheless, in the context of small samples, large effects can be spurious. We (Lopez-Gomez et al. 2019; Lorenzo-Luaces et al. 2017; Lorenzo-Luaces et al. 2020d) and others (Delgadillo and Gonzalez Salas Duhne 2020; Delgadillo et al. 2017; Webb et al. 2020) have used regularization to explore predictors of treatment outcomes.

Power

Statistical power is a function of effect size, sample size, and statistical significance alpha. Setting aside alpha, which is usually selected arbitrarily at $p < 0.05$, a small effect size will make it so that a very large sample is needed to detect statistically significant effects. However, the studies that have cited the multivariable prediction algorithms have varying sample sizes ranging from 84 to 1085. Indeed, while the largest studies had reasonably large samples, the mean size per study arm was 163 and the median was 112. Luedtke et al. (2019) conducted simulations to identify sample sizes that may be needed to detect effects of multivariable treatment assignments under different conditions, manipulating the overall effect of multivariable PTR, the error associated with the PTR, the overall remission rate in the study, and the sample size. According to their results, "at least 300 patients per treatment arm are needed to have adequate statistical power to detect clinically significant underlying marginal improvements" related to the use of a PTR. None of the studies we identified met this criterion fully. No study that we reviewed reported 300 patients per arm, though two came close (Delgadillo and Gonzalez Salas Duhne 2020; Schwartz et al. 2020). Both studies used naturalistic data (i.e., non-randomized treatment assignment). While the use of naturalistic data has been recommended to build multivariable PTRs (Kessler et al. 2019), there are also concerns regarding the use of non-randomized data, even when techniques like propensity score weighting are used to equate differences between treatment groups (DeRubeis 2019).

Schwartz et al. (2020) explored differential treatment outcomes in CBT vs. PDT with a training sample of 966 (CBT 654, PDT 312) and a testing sample of 413 (CBT 281, PDT 132). The effects of the multivariable model on predicting outcome differences between those that had been assigned to their optimal vs. non-optimal treatment in the training data were small but statistically significant ($d = 0.14$, 95% CI = 0.02–

0.27). But, in the testing data, treatment outcome differences between those assigned to their optimal vs. non-optimal treatment were smaller and not statistically significant ($d = 0.09$, 95% CI = −0.11–0.28). Delgadillo and Gonzalez Salas Duhne (2020) explored a model to predict response to CBT for common mental disorder vs. person-centered counseling (CfD). They used a training sample of 1085 patients (CBT $n = 929$ and CfD $n = 156$) and a testing sample of 350 (CBT $n = 175$, CfD $n = 175$). The PTR they developed in a training sample predicted treatment differences between those assigned to their optimal treatment vs. those assigned to their non-optimal treatment. Those that had been assigned to their optimal treatment were more likely to experience reliable and clinically significant depression improvement (62.5%) than those who received their non-optimal treatment (41.7%; OR = 2.33, 95% CI = 1.09, 5.02). While this study was more positive regarding the effects of using PTRs to predict treatment success than the study by Schwartz et al., it is noteworthy that for most patients in the testing sample (68%, $n = 238$), neither treatment was predicted to be superior.

Taken together, these results do not provide strong evidence to believe that multivariable treatment assignment leads to large improvements in patient outcomes, at least in these naturalistic samples. An additional concern regarding the use of naturalistic data to construct PTRs relate to the fidelity with which treatments may be delivered in naturalistic conditions. This includes the degree to which treatment outcomes may be influenced by between-therapist differences, an effect that appears to be larger in naturalistic than RCT data (Lutz et al. 2007; Saxon and Barkham 2012) and which obscures comparisons between treatments.

How Much Can We Expect from PTRs?

Two other avenues of research may be interpreted to suggest that the effects of using multivariable PTRs of response to two treatments of roughly equivalent efficacy may be more limited than usually thought. The first is the line of work on switching therapies for non-responders (Bschor et al. 2018; Rush et al. 2006). This research suggests that while switching treatments sometimes improves outcomes, completing two courses of dissimilar treatments do not lead to remission rates close to 100%. For example, in the Emory PREDICT study, the pooled rate of response after adding CBT or antidepressants to a failed course of either was only 48% (Dunlop et al. 2019). In the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, remission rates in the first stage involving antidepressants were around 37%. Of the subset of patients who agreed to continue treatment, CBT was associated with remission rates of 41% when patients switched and 29% when patients added CBT to their antidepressant (Rush et al. 2006). Similarly, in a recent study of sequential psychological and pharmacological approaches to the treatment of insomnia, Morin et al. (2020) reported remission rates ranging from 36 to 56%, depending on which therapy was initiated. Although, it should be noted, at least one study, of women with recurrent depression, was more optimistic regarding the role of sequential treatment (Frank et al. 2000).

Research on the mechanisms of change in psychotherapy suggests that a sizeable portion of patients who respond in treatment trials will respond to many of the elements of treatment that are common to treatments as well as control conditions, including the passage of time, measurement effects, and expectations of improvement. Cuijpers et al.

(2012) studied the efficacy of non-directive supportive psychotherapy (NDST) for depression, when compared with controls as well as when compared with "specific" therapies like CBT or IPT (see Fig. 1). They used the data on the effects of NDST vs. specific psychotherapies to provide an estimate of the amount of variance in outcomes that is associated with specific therapy techniques above and beyond the general therapeutic setting. They also used the estimates of the efficacy of NDST vs. control conditions as an index of the percentage of variance explained by general conditions common to all psychotherapy (e.g., active listening by a trained professional) vs. the effects provided by various control conditions (used data from a RCT that compared antidepressants e.g., repeated assessment of symptoms, positive experiences). Their calculations (see Fig. 1) suggested that about a third of the change that occurs in treatment trials for depression is accounted for by factors outside of the therapeutic relationship because they are common to waiting lists and other control conditions. About half of the variance was accounted for by non-directive psychosocial support. The remainder, less than a fifth of the variance, was accounted for by specific treatment techniques.

Although the authors caution against the strong interpretation of their estimates and the idea of partitioning "variance explained" in psychotherapy have been criticized (DeRubeis et al. 2014b), these numbers may make one pessimistic about the potential benefits of matching patients, at least to NDST vs. a specific therapeutic condition like
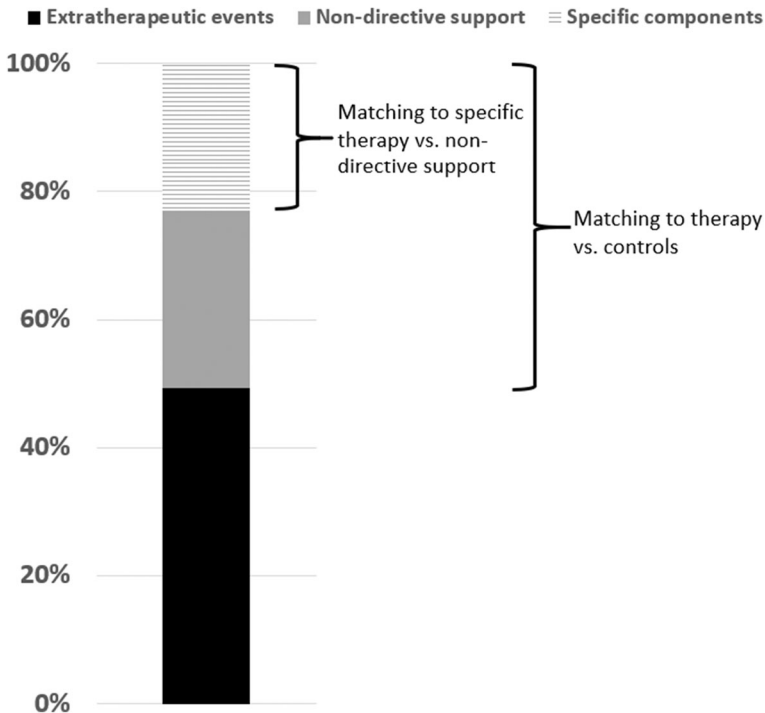
Fig. 1 Variance accounted for by extra-therapeutic actors, presented in control conditions in psychotherapy treatment trials, general therapeutic factors, and treatment-specific psychotherapeutic effects (Cuijpers et al. 2012)

CBT. If only a small percentage of variance is accounted for by specific therapeutic techniques vs. non-specific therapeutic support, it is sensible to question how much could be accounted for contrasts of two different techniques (but, see Serbanescu et al. (2020)). When the techniques are very different (e.g., behavioral activation vs. psychoanalysis), it may be expected that completely different types of patients benefit from the interventions.

An additional cause for concern, coming from research on the efficacy of antidepressants relative to placebos, is the seemingly equal variance in treatment outcomes in antidepressants and placebos (Munkholm et al. 2020; Plöderl and Hengartner 2019). If the ratio of the variance in outcomes in treatment vs. the variance of outcomes in controls is above 1, it may suggest that there are specific subgroups of patients who preferentially respond to treatment. Nonetheless, recent studies suggest that the variance ratio of outcomes in antidepressants vs. placebo is indistinguishable from 1 (Munkholm et al. 2020; Plöderl and Hengartner 2019). While these findings may be taken to suggest that there is limited potential to personalize allocation to antidepressants vs. placebo, and one may be tempted to extend this pessimism to CBT vs. other interventions, much work remains to be done in this topic. For example, Hieronymus et al. (2020) have published data suggesting that the equal variance in antidepressant and placebo outcomes is the product of using study level data means and standard deviations vs. using individual level patient data. Thus, individual patient data meta-analyses should be conducted to test the variance ration in psychotherapy-control studies.

Beyond Depression

We have discussed research on the use of PTRs on depression. However, some have conjectured that depression may be a clinical problems for which many approaches are equally efficacious (Chambless 2002; Westen et al. 2004), possibly reducing the potential for treatment-matching in depression relative to other treatment targets. Our read of the literature suggests that many of the problems we have identified—small samples sizes, small effects, few studies with validation samples, and reasons to doubt the utility of PTRs—are equally present in research on other clinical problems. For example, Niles and colleagues have explored moderators of response in the treatment of anxiety disorders, comparing CBT to acceptance and commitment therapy (ACT) and also comparing coordinated anxiety management with CBT or antidepressants to usual care (Niles et al. 2017a; Niles et al. 2017b). These studies suggest small-to-medium effects matching patients to ACT vs. CBT ($r = 0.28$, $N = 208$) or coordinated anxiety management to usual care ($r = 0.20$, $N = 876$), and they have not been replicated. In the field of substance use treatments, similarly, there has been great interest in the degree of match of patients to therapies. The landmark Project MATCH, which compared CBT to motivational enhancement therapy (MET), and to Twelve-Step Facilitation (TSF, e.g., akin to Alcoholics Anonymous), was a large study ($N = 1726$) testing a priori matching hypothesis in predicting acute and long-term treatment outcomes (Project MATCH Research Group 1998). By and large, the study has been considered a failure in the sense that it neither provided evidence for the superiority of one treatment over another nor did the study find support for any of its matching

hypotheses (Cohen and DeRubeis 2018). While it remains possible that some clinical problems allow more room for treatment-matching than others, it is not clear that there is a specific disorder or symptom for mental for which the promise of PTRs and personalized medicine has been realized.

Idiographic Approaches

We have focused on PTRs derived from data used from comparisons of two treatments. Another approach to personalization is to use idiographic data to personalize the selection of psychotherapy approaches. Fisher and colleagues (Fisher 2015; Fisher and Bosley 2020; Fisher et al. 2019; Fisher and Boswell 2016; Rubel et al. 2018), for example, have used repeated assessments of internalizing symptoms in a period of time prior to the start psychotherapy to identify causal relations between symptoms that exist *within individuals*. This work trades on potential differences in the relations of symptoms between-individuals vs. within-individuals. For example, worry and sadness are highly correlated *between* individuals. Within-individuals, they may also be highly correlated (i.e., some people may be likely to be highly worried at times when they are highly sad), may be uncorrelated, or may be negatively correlated (i.e., when individuals are sad, they may be unlikely to become worried). When expanding across a range of symptoms as well as potential treatment targets (e.g., avoidance), analyzing these idiographic patterns of symptom interrelations may be a data-driven way of assessing, for each patient, what an optimal course of treatment would be. An assumption behind this work is that understanding what symptoms precede and appear causally related to others may yield pathways towards interventions. For example, if sadness, within an individual, triggers a cascade of other symptoms like loss of sleep, substance use, and self-criticism, it would seem like a good target for intervention. Nonetheless, several authors have pointed out that even after estimating longitudinal associations between symptoms (e.g., as in nodes in a network), the causal inferences that can be derived from this work (e.g., symptom X drives symptom Y, decreasing symptom X will decrease symptom Y) are not always straightforward (Bringmann et al. 2019; Dablander and Hinne 2019).

Recommendations

Given the importance of finding large effects for multivariable treatment selection, it becomes incumbent on researchers to identify what it is precisely that can make a treatment-selection endeavor likely to yield large effects. Below, we make some recommendations for future research based on our review of the literature.

Focus on Treatments with Different Mechanisms

DeRubeis et al. (2014b) hypothesized that treatment selection endeavors are most likely to work in a context in which treatments have different underlying mechanisms. For example, in a sample of 291 patients randomized to group CBT vs. a group transdiagnostic CBT (tCBT), Barlow's Unified Protocol (UP), Eskildsen et al. (2020)

examined 36 predictors and moderators of changes in well-being. The authors failed to identify any moderators of response to the treatments. Thus, they could not compute a multivariable index of response to single-disorder CBT vs. tCBT. In hindsight, such an investigation may have been ill-suited to identify moderators of response. An impetus for designing a unified protocol is the idea that single-disorder CBTs share common mechanisms. Transdiagnostic CBT, like "single-disorder" CBTs, engages exposure, along with cognitive reappraisal, to reduce vulnerability to internalizing disorders. It adds a module to enhance motivation, a rationale that emotions are adaptive, and a greater focus on mindfulness awareness of emotions. But there is little evidence that these techniques add to the efficacy of "traditional" CBT techniques or add so much to the intervention so as to make it mechanistically different than single-disorder CBT. Indeed, it should be noted that the UP is a modular intervention so a single course of tCBT with the UP may be very similar to a course of "single-disorder" CBT.

Identifying specific ingredients of psychotherapy has proven to be a major challenge for treatment research (Lorenzo-Luaces et al. 2015; Lorenzo-Luaces et al. 2016). On the one hand, some psychotherapies differ markedly (e.g., exposure and response prevention and psychoanalysis). On the other, some treatments overlap in content considerably (e.g., EMDR, cognitive therapy for PTSD, CPT for PTSD). Even in treatments like IPT, problem-solving therapy (PST), and CBT, there seems to be considerable overlap in mechanisms and outcomes (Lorenzo-Luaces et al. 2015). Thus, it is recommended that researchers attempt multivariable prediction models in contexts where it is expected that there are subgroups that may differ, in part because two treatments may differ. While here we have focused on different mechanisms of efficacy, it may also be possible to focus on mechanisms of treatment delivery. Here too, it is possible that two treatments may have such similar mechanisms (e.g., guided CBT self-help via a book vs. guided CBT self-help via the Internet) that it may not be sensible to expect reliable moderators to emerge from RCT data.

Increasing Heterogeneity in Treatment Trials

Entry criteria (i.e., inclusion and exclusion criteria) are important features of RCTs. Researchers often constrain entry criteria to ensure the validity of their findings, for example, specifically focusing on patients who meet the DSM criteria for MDD. Other entry criteria are frequently used to ensure that the patients who are recruited are only those who would benefit from the level of care provided by the study. For outpatient depression trials, it is common to exclude participants at high suicide risk, those with psychotic symptoms, and individuals who have substance use disorders (SUDs) (Lorenzo-Luaces et al. 2018a, b). For example, in one systematic review, we found that over 85% of treatment trials of psychotherapy or antidepressants for MDD excluded patients on the basis of a SUD. Excluding patients on the basis of a SUD in a trial of outpatient treatments for depression may be an appropriate exclusion for certain studies given doubts about whether depressed mood is substance-induced or questions about the efficacy of depression treatment when substance use disorders are present. Nonetheless, if substance use indeed predicts poor outcomes in treatments for depression, it would be difficult to study this in existing datasets because most excluded participants who used substances at high enough levels to meet DSM criteria for a SUD. One of the reasons the DSM criteria appear largely useless in assigning patients to treatments is that, by and large, treatments studies

have already used as entry criteria DSM symptoms that are associated to prognosis (e.g., psychosis, manic symptoms, suicide risk, symptom severity, substance use). Some of these exclusions are appropriate. For example, there are few reasons to think that a severely catatonic patient would benefit from unguided CBT self-help. In others instances, however, it is not clear that the entry criteria achieve their intended purpose of ensuring the internal validity of the trial. For example, roughly half of individuals who meet criteria for MDD have a substance use disorder at some point in their life (Hasin et al. 2018) so the exclusion is not a trivial one. Nonetheless, the analyses of outcomes in naturalistic settings do not suggest that substance use disorders predict treatment outcomes (Van der Lem et al. 2012). Similarly, while some trials exclude patients on the basis of having severe symptoms, the existing data suggests that depression severity is either not a moderator of treatment outcomes or that more severely ill patients can benefit *more* from active intervention (Bower et al. 2013; Driessen et al. 2010; Fournier et al. 2008; Lorenzo-Luaces et al. 2017, 2020d). Thus, we recommend that researchers be more intentional about their entry criteria, only excluding patients when they have strong reasons to believe they will not benefit from treatment as well as disinterest in the status of the entry variable as a potential moderator.

Biomarkers

There has been a great deal of interest in the use of biomarkers, especially genetic variables and neuroimaging findings to match patients to treatments (Thase 2014). In one study, for example, McGrath et al. (2013) used positron emission tomography (PET) to identify baseline activity in brain areas that were associated with response to CBT vs. antidepressants ($N = 38$). They identified several areas that were potential moderators of response, including the right anterior insula which appeared to be the area most strongly associated with differential response (SMD = 1.43). Large multisite studies have attempted to use biomarkers, including by identifying depression subtypes ($n = 711$) that may predict treatment response to transcranial magnetic stimulation ($n = 154$) (Drysdale et al. 2017) or using electroencephalography to identify responders to antidepressants vs. placebos (Fatt et al. 2020). Because these endeavors still task researchers with identifying interactions between pre-treatment variables (e.g., a specific genetic polymorphism) and treatment in predicting outcomes, large samples and rigorous analytic methods are still needed. For example, Dinga et al. (2019) criticized the analysis by Drysdale et al. for its proneness to overfitting. They replicated the analysis of Drysdale et al. but provided evidence that spuriousness could not be ruled out as causing the results.

Thus, many of the issues identifying on research on PTRs also apply to biomarkers. Biomarkers may prove very advantageous to use if they are subject to less measurement error relative to other variables typically used to measure potential pre-treatment moderators because measurement error is another factor that may complicate detection of a PTR (Luedtke et al. 2019).

Novel Measurements

One of the considerations when developing PTRs to allocate patients to treatments is which variables make up the decision rule. Virtually, all of the studies that explored

moderators explored self-reported questionnaires or rating scales completed by observers. An issue with these measurements is that they are often subjective, may be correlated due to monomethod bias, and may be resource-intensive to administer. Most (90%) Americans have Internet access, and more than three-quarters of Americans own a smartphone (Pew Research Center: Internet and Technology 2019). Advances in digital technologies provide new opportunities to increase treatment and assessment access to many as well as to customize interventions. For example, TestMyBrain.org is a citizen science initiative where participants take part in research experiments to contribute to science and learn about themselves through immediate and personalized assessments that return research results. Extending far outside traditional research and clinical settings, TestMyBrain and other digital research laboratories provide benefits to patients and generate scientific insights that would not have been possible with traditional models.

A growing literature shows that tests administered on the web are comparable in quality to tests administered in research labs (Germine et al. 2012; Hartshorne and Germine 2015; Meyerson and Tryon 2003). In addition to web-based tests, social media represents a unique resource of publicly available, real-time data that can help characterize the trajectory of mental illness as well as behavioral and environmental risk factors at unprecedented resolution and samples. Social media is ideally suited for characterizing risk trajectories for mental illness because multiple dimensions of data, individual and social, can be collected on large samples and across multiple data points. Twitter, for example, is used daily by over a fifth of all people in the USA (Pew Research Center: Internet and Technology 2019). We have shown that social media data can be mined to characterize the cognitive (Bathina et al. 2020) as well as behavioral indicators of circadian rhythm (ten Thij et al. 2020). Further consideration of variables that can be captured through passive sensing, analysis of social media data and cognitive tests administered remotely have the potential to further enrich PTRs.

Summary and Conclusions

No single psychotherapy for depression works for all patients or even is superior to other treatments (Barth et al. 2016). Accordingly, researchers are currently exploring individual differences that may predict response to CBTs and other treatments (Cohen and DeRubeis 2018). Some of the usual suspects, including demographics and treatment preferences, do not appear to be reliable predictors or moderators. Likewise, the DSM symptoms of MDD do not appear to be reliable moderators of outcomes (Boschloo et al. 2019; Kappelmann et al. 2020). Illness history features, like severity and chronicity, may be better predictors of outcomes as are symptoms that are already used for stratification like psychosis and manic symptoms. Similarly, there is some evidence that capitalizing on existing strengths that patients have leads to better outcomes in therapy (Cheavens et al. 2012). While risk stratification by illness features and compensating on patients' relative weaknesses may be promising lines of inquiry, researchers are increasingly using data-driven approaches to identify multiple moderators of outcomes.

We reviewed the literature that has been the product of attempts to replicate studies in which atheoretical multivariable prediction algorithm models were developed. By

and large, there are a wide variety of methods (Barber and Muenz 1996; Beutler et al. 1991; Cloitre et al. 2016; DeRubeis et al. 2014a; Kraemer 2013; Lorenzo-Luaces et al. 2017; Lutz et al. 2006; Wallace et al. 2013) that have been used to combine potential moderators of outcomes to CBTs and other treatment(s). Accordingly, there has been substantial variability in the reporting of metrics that characterize the supposed benefit of matching patients to treatments. This literature is characterized by re-analyses of small RCTs in which the effects that have been reported are relatively modest. Similar critiques have been made by Kessler and colleagues (Luedtke et al. 2019) who have observed that very large ($N > 600$) trials are needed to have adequate power to detect the types of effects researchers may find when using multivariable algorithms. To make matters worse, there have been only a handful of attempts to test models out of the sample in which they have been developed (Delgadillo and Gonzalez Salas Duhne 2020; Delgadillo et al. 2017; Schwartz et al. 2020), and these studies do not paint a compelling picture regarding the promise of using multivariable PTRs to assign patients to treatments.

Researchers have largely tried to predict response to psychotherapies of different modalities, though the existing evidence suggests that a large portion of the variance in average treatment outcomes is best explained by factors that are common across evaluation settings and across psychosocial interventions. Accordingly, it may be more promising to conduct studies to match patients to higher vs. lower levels of care, for example, unguided self-help vs. face-to-face CBT (Kessler 2018). Given that low-intensity CBT options tend to be less costly to implement, studying these interventions allows researchers to collect data on larger samples of individuals, which will become a necessity to develop reliable PTRs.

Compliance with Ethical Standards

References

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2020). A scoping review of machine learning in psychotherapy research. *Psychother Res*, 1–25. Advance online publication. https://doi.org/10.1080/10503307.2020.1808729.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. American Psychiatric Association: Washington, DC.

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*(5), 951–958. https://doi.org/10.1037/0022-006X.64.5.951.

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. *PLOS Medicine, 10*, e1001454. https://doi.org/10.1371/journal.pmed.1001454.

Bathina, K. C., Thij, M. t., Lorenzo-Luaces, L., Rutter, L. A., & Bollen, J. (2020). Depressed individuals express more distorted thinking on social media. *arXiv preprint arXiv:2002.02800*.

Baumeister, H., & Parker, G. (2012). Meta-review of depressive subtyping models. *J Affect Disord, 139*(2), 126–140. https://doi.org/10.1016/j.jad.2011.07.015.

Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., & Merry, W. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *J Consult Clin Psychol, 59*(2), 333–340. https://doi.org/10.1037/0022-006X.59.2.333.

Bockting, C. L., Hollon, S. D., Jarrett, R. B., Kuyken, W., & Dobson, K. (2015). A lifetime approach to major depressive disorder: the contributions of psychological interventions in preventing relapse and recurrence. *Clin Psychol Rev, 41*, 16–26. https://doi.org/10.1016/j.cpr.2015.02.003.

Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S., Dunner, D. L., Dunlop, B. W., Hegerl, U., Hollon, S. D., et al. (2019). The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression: results from an individual patient data meta-analysis. *World Psychiatry, 18*(2), 183–191. https://doi.org/10.1002/wps.20630.

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., et al. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *Br Med J, 346*, f540. https://doi.org/10.1136/bmj.f540.

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J. T., & Snippe, E. (2019). What do centrality measures measure in psychological networks? *J Abnorm Psychol, 128*(8), 892.

Bruijniks, S. J., Lemmens, L. H., Hollon, S. D., Peeters, F. P., Cuijpers, P., Arntz, A., Dingemanse, P., Willems, L., Van Oppen, P., Twisk, J. W., et al. (2020). The effects of once-versus twice-weekly sessions on psychotherapy outcomes in depressed patients. *Br J Psychiatry, 216*(4), 222–230. https://doi.org/10.1192/bjp.2019.265.

Bschor, T., Kern, H., Henssler, J., & Baethge, C. (2018). Switching the antidepressant after nonresponse in adults with major depression: a systematic literature search and meta-analysis., *79*(1), 16r10749. https://doi.org/10.4088/JCP.16r10749.

Chambless, D. L. (2002). Beware the dodo bird: the dangers of overgeneralization. *Clin Psychol Sci Pract, 9*(1), 13–16. https://doi.org/10.1093/clipsy.9.1.13.

Cheavens, J. S., Strunk, D. R., Lazarus, S. A., & Goldstein, L. A. (2012). The compensation and capitalization models: a test of two approaches to individualizing the treatment of depression. *Behav Res Ther, 50*(11), 699–706. https://doi.org/10.1016/j.brat.2012.08.002.

Cloitre, M., Petkova, E., Su, Z., & Weiss, B. J. (2016). Patient characteristics as a moderator of posttraumatic stress disorder treatment outcome: combining symptom burden and strengths. *BJPsych Open, 2*(2), 101–106. https://doi.org/10.1192/bjpo.bp.115.000745.

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.

Cohen, D., Deniau, E., Maturana, A., Tanguy, M.-L., Bodeau, N., Labelle, R., Breton, J.-J., & Guile, J.-M. (2008). Are child and adolescent responses to placebo higher in major depression than in anxiety disorders? A systematic review of placebo-controlled trials. *PLoS One, 3*(7), e2632. https://doi.org/10.1371/journal.pone.0002632.

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2020). A demonstration of a multi-method variable selection approach for treatment selection: recommending cognitive–behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychother Res, 30*(2), 137–150. https://doi.org/10.1080/10503307.2018.1563312.

Cuijpers, P., van Straten, A., Schuurmans, J., van Oppen, P., Hollon, S. D., & Andersson, G. (2010). Psychotherapy for chronic major depression and dysthymia: a meta-analysis. *Clin Psychol Rev, 30*(1), 51–62. https://doi.org/10.1016/j.cpr.2009.09.003.

Cuijpers, P., Driessen, E., Hollon, S. D., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: a meta-analysis. *Clin Psychol Rev, 32*(4), 280–291. https://doi.org/10.1016/j.cpr.2012.01.003.

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

Cuijpers, P., Weitz, E., Twisk, J., Kuehner, C., Cristea, I., David, D., DeRubeis, R. J., Dimidjian, S., Dunlop, B. W., Faramarzi, M., et al. (2014). Gender as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression: an "individual patient data" meta-analysis. *Depression and Anxiety, 31*(11), 941–951. https://doi.org/10.1002/da.22328.

Cuijpers, P., Weitz, E., Lamers, F., Penninx, B. W., Twisk, J., DeRubeis, R. J., Dimidjian, S., Dunlop, B. W., Jarrett, R. B., Segal, Z. V., et al. (2017). Melancholic and atypical depression as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression. *Depression and Anxiety, 34*(3), 246–256. https://doi.org/10.1002/da.22580.

Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., & Furukawa, T. A. (2019). Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. *JAMA Psychiatry, 76*(7), 700–707. https://doi.org/10.1001/jamapsychiatry.2019.0268.

Cuijpers, P., Karyotaki, E., Eckshtain, D., Ng, M. Y., Corteselli, K. A., Noma, H., Quero, S., & Weisz, J. R. (2020). Psychotherapy for depression across different age groups: a systematic review and meta-analysis. *JAMA Psychiatry, 77*(7), 694–702. https://doi.org/10.1001/jamapsychiatry.2020.0164.

Dablander, F., & Hinne, M. (2019). Node centrality measures are a poor substitute for causal inference. *Sci Rep, 9*(1), 1–13.

Delgadillo, J., & Gonzalez Salas Duhne, P. (2020). Targeted prescription of cognitive–behavioral therapy versus person-centered counseling for depression using a machine learning approach. *Journal of Consulting and Clinical Psychology, 88*(1), 14–24. https://doi.org/10.1037/ccp0000476.

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *J Consult Clin Psychol, 85*(9), 835–853. https://doi.org/10.1037/ccp0000231.

DeRubeis, R. J. (2019). The history, current status, and possible future of precision mental health. *Behav Res Ther, 123*, 103506. https://doi.org/10.1016/j.brat.2019.103506.

DeRubeis, R. J., & Lorenzo-Luaces, L. (2017). Recognizing that truth is unattainable and attending to the most informative research evidence. *Psychother Res, 27*(1), 33–35. https://doi.org/10.1080/10503307.2016.1252072.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., O'Reardon, J. P., Lovett, M. L., Gladis, M. M., Brown, L. L., et al. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry, 62*(4), 409–416. https://doi.org/10.1001/archpsyc.62.4.409.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014a). 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., Gelfand, L. A., German, R. E., Fournier, J. C., & Forand, N. R. (2014b). Understanding processes of change: how some patients reveal more than others—and some groups of therapists less—about what matters in psychotherapy. *Psychother Res, 24*(3), 419–428. https://doi.org/10.1080/10503307.2013.838654.

Dinga, R., Schmaal, L., Penninx, B. W., van Tol, M. J., Veltman, D. J., van Velzen, L., Mennes, M., van der Wee, N. J., & Marquand, A. F. (2019). Evaluating the evidence for biotypes of depression: methodological replication and extension of Drysdale et al. (2017). *NeuroImage: Clinical, 22*, 101796. https://doi.org/10.1016/j.nicl.2019.101796.

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. *J Consult Clin Psychol, 78*(5), 668–680. https://doi.org/10.1037/a0020570.

Driessen, E., Van, H. L., Don, F. J., Peen, J., Kool, S., Westra, D., Hendriksen, M., Schoevers, R. A., Cuijpers, P., Twisk, J. W., et al. (2013). The efficacy of cognitive-behavioral therapy and psychodynamic therapy in the outpatient treatment of major depression: a randomized clinical trial. *Am J Psychiatr, 170*(9), 1041–1050. https://doi.org/10.1176/appi.ajp.2013.12070899.

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R. N., Zebley, B., Oathes, D. J., Etkin, A., et al. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med, 23*(1), 28–38. https://doi.org/10.1038/nm.4246.

Dunlop, B. W., LoParo, D., Kinkead, B., Mletzko-Crowe, T., Cole, S. P., Nemeroff, C. B., Mayberg, H. S., & Craighead, W. E. (2019). Benefits of sequentially adding cognitive-behavioral therapy or antidepressant medication for adults with nonremitting depression. *Am J Psychiatr, 176*(4), 275–286. https://doi.org/10.1176/appi.ajp.2018.18091075.

Eskildsen, A., Reinholt, N., van Bronswijk, S. C., Brund, R. B., Christensen, A. B., Hvenegaard, M., Arendt, M., Alro, A., Poulsen, S., Rosenberg, N. K., et al. (2020). Personalized psychotherapy for outpatients with major depression and anxiety disorders: transdiagnostic versus diagnosis-specific group cognitive behavioural therapy. *Cogn Ther Res, 44*, 988–1001. https://doi.org/10.1007/s10608-020-10116-1.

Fatt, C. R. C., Cooper, C., Jha, M. K., Aslan, S., Grannemann, B., Kurian, B., Greer, T. L., Fava, M., Weissman, M., McGrath, P. J., et al. (2020). Dorsolateral prefrontal cortex and subcallosal cingulate

connectivity show preferential antidepressant response in major depression. *Biol Psychiatry*. Advance online publication. https://doi.org/10.1016/j.bpsc.2020.06.019.

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: implications for personalized care. *J Consult Clin Psychol, 83*(4), 825–836. https://doi.org/10.1037/ccp0000026.

Fisher, A. J., & Bosley, H. G. (2020). Identifying the presence and timing of discrete mood states prior to therapy. *Behaviour Research and Therapy, 128*, 103596. https://doi.org/10.1016/j.brat.2020.103596.

Fisher, A. J., & Boswell, J. F. (2016). Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment, 23*(4), 496–506. https://doi.org/10.1177/1073191116638735.

Fisher, A. J., Bosley, H. G., Fernandez, K. C., Reeves, J. W., Soyster, P. D., Diamond, A. E., & Barkin, J. (2019). Open trial of a personalized modular treatment for mood and anxiety. *Behav Res Ther, 116*, 69–79. https://doi.org/10.1016/j.brat.2019.01.010.

Flückiger, C., Forrer, L., Schnider, B., Bättig, I., Bodenmann, G., & Zinbarg, R. E. (2016). A single-blinded, randomized clinical trial of how to implement an evidence-based treatment for generalized anxiety disorder [IMPLEMENT]—effects of three different strategies of implementation. *EBio Medicine, 3*, 163–171. https://doi.org/10.1016/j.ebiom.2015.11.049.

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*(2), 124–129. https://doi.org/10.1192/bjp.bp.107.037234.

Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Hollon, S. D., Amsterdam, J. D., & Gallop, R. (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *J Consult Clin Psychol, 77*(4), 775–787. https://doi.org/10.1037/a0015401.

Frank, E., Grochocinski, V. J., Spanier, C. A., Buysse, D. J., Cherry, C. R., Houck, P. R., Stapf, D. M., & Kupfer, D. J. (2000). Interpersonal psychotherapy and antidepressant medication: evaluation of a sequential treatment strategy in women with recurrent major depression. *The Journal of clinical psychiatry, 1*, 51–57.

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

Fried, E. I., Coomans, F., & Lorenzo-Luaces, L. (2020). The 341737 ways of qualifying for the melancholic specifier. *The Lancet Psychiatry, 7*(6), 479–480. https://doi.org/10.1016/S2215-0366(20)30169-3.

Friedl, N., Berger, T., Krieger, T., Caspar, F., & Grosse Holtforth, M. (2020). Using the personalized advantage index for individual treatment allocation to cognitive behavioral therapy (CBT) or a cbt with integrated exposure and emotion-focused elements (CBT-EE). *Psychotherapy Research, 30*(6), 763–775. https://doi.org/10.1080/10503307.2019.1664782.

Gaspersz, R., Lamers, F., Kent, J. M., Beekman, A. T., Smit, J. H., van Hemert, A. M., Schoevers, R. A., & Penninx, B. W. (2017). Anxious distress predicts subsequent treatment outcome and side effects in depressed patients starting antidepressant treatment. *Journal of Psychiatric Research, 84*, 41–48. https://doi.org/10.1016/j.jpsychires.2016.09.018.

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychon Bull Rev, 19*(5), 847–857. https://doi.org/10.3758/s13423-012-0296-9.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychol Sci, 26*(4), 433–443. https://doi.org/10.1177/0956797614567339.

Hasin, D. S., Sarvet, A. L., Meyers, J. L., Saha, T. D., Ruan, W. J., Stohl, M., & Grant, B. F. (2018). Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry, 75*(4), 336–346. https://doi.org/10.1001/jamapsychiatry.2017.4602.

Hieronymus, F., Hieronymus, M., Nilsson, S., Eriksson, E., & Østergaard, S. D. (2020). Individual variability in treatment response to antidepressants in major depression: comparing trial-level and patient-level analyses. *Acta Psychiatr Scand*. Accepted Author Manuscript. https://doi.org/10.1111/acps.13205.

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: a randomized clinical trial. *JAMA Psychiatry, 71*(10), 1157–1164. https://doi.org/10.1001/jamapsychiatry.2014.1054.

Hollon, S. D., Cohen, Z. D., Singla, D. R., & Andrews, P. W. (2019). Recent developments in the treatment of depression. *Behav Ther, 50*(2), 257–269. https://doi.org/10.1016/j.beth.2019.01.002.

Huibers, M., Cohen, Z. D., Lemmens, L., HJM, L., Arntz, A., Peeters, F., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed

individuals using the personalized advantage index approach. *PLoS One, 10*(11), e0140771. https://doi.org/10.1371/journal.pone.0140771.

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. *The British Journal of Psychiatry, 190*(6), 503–508. https://doi.org/10.1192/bjp.bp.106.024737.

Kappelmann, N., Rein, M., Fietz, J., Mayberg, H. S., Craighead, W. E., Dunlop, B. W., Nemeroff, C. B., Keller, M., Klein, D. N., Arnow, B. A., et al. (2020). Psychotherapy or medication for depression? Using individual symptom meta-analyses to derive a symptom-oriented therapy (SOrT) metric for a personalised psychiatry. *BMC Med, 18*, 170. https://doi.org/10.1186/s12916-020-01623-9.

Karyotaki, E., Ebert, D. D., Donkin, L., Riper, H., Twisk, J., Burger, S., Rozental, A., Lange, A., Williams, A. D., Zarski, A. C., et al. (2018). Do guided Internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. *Clin Psychol Rev, 63*, 80–92. https://doi.org/10.1016/j.cpr.2018.06.007.

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

Kessler, R. C., Bossarte, R. M., Luedtke, A., Zaslavsky, A. M., & Zubizarreta, J. R. (2019). Machine learning methods for developing precision treatment rules with observational data. *Behav Res Ther, 120*, 103412. https://doi.org/10.1016/j.brat.2019.103412.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., et al. (2017). The hierarchical taxonomy of psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *J Abnorm Psychol, 126*(4), 454. https://doi.org/10.1037/abn0000258.

Kraemer, H. C. (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. *Stat Med, 32*(11), 1964–1973. https://doi.org/10.1002/sim.5734.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry, 59*(10), 877–883. https://doi.org/10.1001/archpsyc.59.10.877.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci, 12*(5), 535–540. https://doi.org/10.1038/nn.2303.

Lee, Y., Ragguett, R.-M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A., Brietzke, E., Lin, K., Pan, Z., Subramaniapillai, M., et al. (2018). Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *J Affect Disord, 241*, 519–532. https://doi.org/10.1016/j.jad.2018.08.073.

Lemmens, L. H., van Bronswijk, S. C., Peeters, F., Arntz, A., Hollon, S. D., & Huibers, M. J. (2019). Long-term outcomes of acute treatment with cognitive therapy v. interpersonal psychotherapy for adult depression: Follow-up of a randomized controlled trial. *Psychol Med, 49*(3), 465–473. https://doi.org/10.1017/S0033291718001083.

Leykin, Y., Amsterdam, J. D., DeRubeis, R. J., Gallop, R., Shelton, R. C., & Hollon, S. D. (2007). Progressive resistance to a selective serotonin reuptake inhibitor but not to cognitive therapy in the treatment of major depression. *J Consult Clin Psychol, 75*(2), 267–276. https://doi.org/10.1037/0022-006X.75.2.267.

Locher, C., Koechlin, H., Zion, S. R., Werner, C., Pine, D. S., Kirsch, I., Kessler, R. C., & Kossowsky, J. (2017). Efficacy and safety of selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, and placebo for common psychiatric disorders among children and adolescents: a systematic review and meta-analysis. *JAMA Psychiatry, 74*(10), 1011–1020. https://doi.org/10.1001/jamapsychiatry.2017.2432.

Lopez-Gomez, I., Lorenzo-Luaces, L., Chaves, C., Hervas, G., DeRubeis, R. J., & Vazquez, C. (2019). Predicting optimal interventions for clinical depression: moderators of outcomes in a positive psychological intervention vs. cognitive-behavioral therapy. *Gen Hosp Psychiatry, 61*, 104–110. https://doi.org/10.1016/j.genhosppsych.2019.07.004.

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., German, R. E., & DeRubeis, R. J. (2015). It's complicated: the relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clin Psychol Rev, 41*, 3–15. https://doi.org/10.1016/j.cpr.2014.12.003.

Lorenzo-Luaces, L., Keefe, J. R., & DeRubeis, R. J. (2016). Cognitive-behavioral therapy: nature and relation to non-cognitive behavioral therapy. *Behav Ther, 47*(6), 785–803. https://doi.org/10.1016/j.beth.2016.02.012.

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. *J Affect Disord, 213*, 78–85. https://doi.org/10.1016/j.jad.2017.02.010.

Lorenzo-Luaces, L., Johns, E., & Keefe, J. R. (2018a). The generalizability of randomized controlled trials of self-guided Internet-based cognitive behavioral therapy for depressive symptoms: systematic review and meta-regression analysis. *J Med Internet Res, 20*(11), e10113. https://doi.org/10.2196/10113.

Lorenzo-Luaces, L., Zimmerman, M., & Cuijpers, P. (2018b). Are studies of psychotherapies for depression more or less generalizable than studies of antidepressants? *J Affect Disord, 234*, 8–13. https://doi.org/10.1016/j.jad.2018.02.066.

Lorenzo-Luaces, L., Buss, J., & Fried, E. I. (2020a). *Do melancholic and atypical specifiers reduce heterogeneity in major depressive disorder?: A secondary analysis of the STAR*D trial*. Manuscript submitted for publication. https://doi.org/10.31234/osf.io/ubndc.

Lorenzo-Luaces, L., Lemmens, L. H., Keefe, J. R., Cuijpers, P., & Bockting, C. L. (2020b). The efficacy of cognitive behavioral therapy for depression and anxiety: an updated review of meta-analyses (A. Wenzel, Ed.). In A. Wenzel (Ed.), *Handbook of cognitive behavioral therapy*. Oxford: Oxford University Press.

Lorenzo-Luaces, L., Rodriguez-Quintana, N., & Bailey, A. J. (2020c). Double trouble: Do depression severity and duration interact to predicting treatment outcomes in adolescent depression? *Behav Res Ther, 131*, 103637. https://doi.org/10.1016/j.brat.2020.103637.

Lorenzo-Luaces, L., Rodriguez-Quintana, N., Riley, T. N., & Weisz, J. R. (2020d). 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*. Advance online publication. https://doi.org/10.1080/10503307.2020.1747657.

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science, 7*(3), 445–461. https://doi.org/10.1177/2167702618815466.

Lutz, W., Saunders, S. M., Leon, S. C., Martinovich, Z., Kosfelder, J., Schulte, D., Grawe, K., & Tholen, S. (2006). Empirically and clinically useful decision making in psychotherapy: differential predictions with treatment response models. *Psychol Assess, 18*(2), 133–141. https://doi.org/10.1037/1040-3590.18.2.133.

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: a three-level growth curve approach. *J Couns Psychol, 54*(1), 32. https://doi.org/10.1037/0022-0167.54.1.32.

McBride, C., Atkinson, L., Quilty, L. C., & Bagby, R. M. (2006). Attachment as moderator of treatment outcome in major depression: a randomized control trial of interpersonal psychotherapy versus cognitive behavior therapy. *J Consult Clin Psychol, 74*(6), 1041. https://doi.org/10.1037/0022-006X.74.6.1041.

McGrath, C. L., Kelley, M. E., Holtzheimer, P. E., Dunlop, B. W., Craighead, W. E., Franco, A. R., Craddock, R. C., & Mayberg, H. S. (2013). Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry, 70*(8), 821–829. https://doi.org/10.1001/jamapsychiatry.2013.143.

Meyerson, P., & Tryon, W. W. (2003). Validating Internet research: a test of the psychometric equivalence of Internet and in-person samples. *Behav Res Methods Instrum Comput, 35*(4), 614–620. https://doi.org/10.3758/BF03195541.

Monroe, S. M., & Harkness, K. L. (2005). Life stress, the "kindling" hypothesis, and the recurrence of depression: considerations from a life stress perspective. *Psychol Rev, 112*(2), 417–445. https://doi.org/10.1037/0033-295X.112.2.417.

Monroe, S. M., & Harkness, K. L. (2011). Recurrence in major depression: a conceptual analysis. *Psychol Rev, 118*(4), 655–674. https://doi.org/10.1037/a0025190.

Morin, C. M., et al. (2020). Effectiveness of sequential psychological and medication therapies for insomnia disorder: a randomized clinical trial. *JAMA Psychiatry. 77*(11), 1107–1115. https://doi.org/10.1001/jamapsychiatry.2020.1767.

Munkholm, K., Winkelbeiner, S., & Homan, P. (2020). Individual response to antidepressants for depression in adults-a meta-analysis and simulation study. *PLoS One, 15*(8), e0237950. https://doi.org/10.1371/journal.pone.0237950.

Niles, A. N., Loerinc, A. G., Krull, J. L., Roy-Byrne, P., Sullivan, G., Sherbourne, C. D., Bystritsky, A., & Craske, M. G. (2017a). Advancing personalized medicine: application of a novel statistical method to identify treatment moderators in the coordinated anxiety learning and management study. *Behav Ther, 48*(4), 490–500. https://doi.org/10.1016/j.beth.2017.02.001.

Niles, A. N., Wolitzky-Taylor, K. B., Arch, J. J., & Craske, M. G. (2017b). Applying a novel statistical method to advance the personalized treatment of anxiety disorders: a composite moderator of comparative drop-out from CBT and ACT. *Behav Res Ther, 91*, 13–23. https://doi.org/10.1016/j.brat.2017.01.001.

Olbert, C. M., Gala, G. J., & Tupler, L. A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. *J Abnorm Psychol, 123*(2), 452–462. https://doi.org/10.1037/a0036068.

Otero, P., Smit, F., Cuijpers, P., DeRubeis, R. J., & Torres, A´., & Va´zquez, F. L. (2015). Differential response to depression prevention among a sample of informal caregivers: moderator analysis of longer-term follow-up trial data. *Psychiatry Res, 230*(2), 271–278. https://doi.org/10.1016/j.psychres.2015.09.005.

Pae, C.-U., Vöhringer, P. A., Holtzman, N. S., Thommi, S. B., Patkar, A., Gilmer, W., & Ghaemi, S. N. (2012). Mixed depression: a study of its phenomenology and relation to treatment response. *Journal of Affective Disorders, 136*(3), 1059–1061. https://doi.org/10.1016/j.jad.2011.11.024.

Patel, V., Chisholm, D., Parikh, R., Charlson, F. J., Degenhardt, L., Dua, T., Ferrari, A. J., Hyman, S., Laxminarayan, R., Levin, C., et al. (2016). Addressing the burden of mental, neurological, and substance use disorders: key messages from disease control priorities. *The Lancet, 387*(10028), 1672–1685. https://doi.org/10.1016/S0140-6736(15)00390-6.

Perez, J., Russo, D. A., Stochl, J., Clarke, J., Martin, Z., Jassi, C., French, P., Fowler, D., & Jones, P. B. (2018). Common mental disorder including psychotic experiences: trailblazing a new recovery pathway within the improving access to psychological therapies programme in England. *Early Intervention in Psychiatry, 12*(3), 497–504. https://doi.org/10.1111/eip.12434.

Petkova, E., Park, H., Ciarleglio, A., Ogden, R. T., & Tarpey, T. (2020). Optimising treatment decision rules through generated effect modifiers: a precision medicine tutorial. *BJPsych Open, 6*(1), E2. https://doi.org/10.1192/bjo.2019.85.

Pew Research Center: Internet & Technology. (2019). Social media fact sheet. Retrieved September 8, 2020, from https://www.pewresearch.org/internet/fact-sheet/social-media/.

Plöderl, M., & Hengartner, M. P. (2019). What are the chances for personalised treatment with antidepressants? Detection of patient-by-treatment interaction with a variance ratio meta-analysis. *BMJ open, 9*, e034816. https://doi.org/10.1136/bmjopen-2019-034816.

Polo, A. J., Makol, B. A., Castro, A. S., Col'on-Quintana, N., Wagstaff, A. E., & Guo, S. (2019). Diversity in randomized clinical trials of depression: a 36-year review. *Clin Psychol Rev, 67*, 22–35. https://doi.org/10.1016/j.cpr.2018.09.004.

Project MATCH Research Group. (1998). Matching alcoholism treatments to client heterogeneity: project MATCH three-year drinking outcomes. *Alcohol Clin Exp Res, 22*(6), 1300–1311. https://doi.org/10.1111/j.1530-0277.1998.tb03912.x.

Rekkas, A., Paulus, J. K., Raman, G., Wong, J. B., Steyerberg, E. W., Rijnbeek, P. R., Kent, D. M., & van Klaveren, D. (2020). Predictive approaches to heterogeneous treatment effects: a scoping review. *BMC Med Res Methodol* 20, 264. https://doi.org/10.1186/s12874-020-01145-1.

Rubel, J. A., Fisher, A. J., Husen, K., & Lutz, W. (2018). Translating person-specific network models into personalized treatments: development and demonstration of the dynamic assessment treatment algorithm for individual networks (DATA-IN). *Psychother Psychosom, 87*(4), 249–251. https://doi.org/10.1159/000487769.

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., et al. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *Am J Psychiatr, 163*(11), 1905–1917.

Sasso, K. E., Strunk, D. R., Braun, J. D., DeRubeis, R. J., & Brotman, M. A. (2015). Identifying moderators of the adherence-outcome relation in cognitive therapy for depression. *J Consult Clin Psychol, 83*(5), 976–984. https://doi.org/10.1037/ccp0000045.

Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: therapist effects and the contribution of patient severity and risk. *J Consult Clin Psychol, 80*(4), 535–546. https://doi.org/10.1037/a0028898.

Schwartz, B., Cohen, Z. D., Rubel, J. A., Zimmermann, D., Wittmann, W. W., & Lutz, W. (2020). Personalized treatment selection in routine care: integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research* https://doi.org/10.1080/10503307.2020.1769219.

Serbanescu, I., Walter, H., Schnell, K., Kessler, H., Weber, B., Drost, S., Groß, M., Neudeck, P., Klein, J. P., Assmann, N., Zobel, I., Backenstrass, M., Hautzinger, M., Meister, R., & Härter, M., Schramm, E., & Schoepf, D. (2020). Combining baseline characteristics to disentangle response differences to disorder-specific versus supportive psychotherapy in patients with persistent depressive disorder. *Behav Res Ther, 124*, 103512. https://doi.org/10.1016/j.brat.2019.103512.

ten Thij, M., van de Leemput, I., Bathina, K., Rutter, L., Lorenzo-Luaces, L., Scheffer, M., & Bollen, J. (2020). Depression alters circadian pattern of online activity. *Sci Rep, 10*, 17272. https://doi.org/10.1038/s41598-020-74314-3.

Thase, M. E. (2014). Using biomarkers to predict treatment response in major depressive disorder: evidence from past and present studies. *Dialogues Clin Neurosci, 16*(4), 539.

Ungvari, G. S., Kau, L. S., Wai-Kw, T., & Shing, N. F. (2001). The pharmacological treatment of catatonia: an overview. *Eur Arch Psychiatry Clin Neurosci, 251*(1), I31–I34. https://doi.org/10.1007/PL00014198.

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? *Psychological medicine.* https://doi.org/10.1017/S0033291719003192.

van Bronswijk, S. C., Sanne J E Bruijniks, Lorenzo-Luaces, L., DeRubeis, R. J., Lemmens, L. H. J. M., Peeters, F. P., & Huibers, M. J. (2020). *Cross-trial prediction in psychotherapy: external validation of the Personalized Advantage Index using machine learning in two Dutch randomized trials comparing CBT versus IPT for depression.* Manuscript submitted for publication.

Van der Lem, R., De Wever, W. W., Van der Wee, N. J., Van Veen, T., Cuijpers, P., & Zitman, F. G. (2012). The generalizability of psychotherapy efficacy trials in major depressive disorder: an analysis of the influence of patient selection in efficacy trials on symptom outcome in daily practice. *BMC Psychiatry, 12*, 192. https://doi.org/10.1186/1471-244X-12-192.

Wallace, M. L., Frank, E., & Kraemer, H. C. (2013). A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *JAMA Psychiatry, 70*(11), 1241–1247. https://doi.org/10.1001/jamapsychiatry.2013.1960.

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

Weitz, E. S., Hollon, S. D., Twisk, J., Van Straten, A., Huibers, M. J., David, D., DeRubeis, R. J., Dimidjian, S., Dunlop, B. W., Cristea, I. A., et al. (2015). Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: an individual patient data meta-analysis. *JAMA Psychiatry, 72*(11), 1102–1109. https://doi.org/10.1001/jamapsychiatry.2015.1516.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychol Bull, 130*(4), 631. https://doi.org/10.1037/0033-2909.130.4.631.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., et al. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. *Lancet, 382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6.

Windle, E., Tee, H., Sabitova, A., Jovanovic, N., Priebe, S., & Carr, C. (2020). Association of patient treatment preference with dropout and clinical outcomes in adult psychosocial mental health interventions: a systematic review and meta-analysis. *JAMA Psychiatry, 77*(3), 294–302. https://doi.org/10.1001/jamapsychiatry.2019.3750.

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Compr Psychiatry, 56*, 29–34. https://doi.org/10.1016/j.comppsych.2014.09.007.