ANNUAL REVIEWS

Annual Review of Clinical Psychology

Treatment Selection in Depression

Zachary D. Cohen and Robert J. DeRubeis

Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: zachary.d.cohen@gmail.com

Annu. Rev. Clin. Psychol. 2018. 14:209–36

First published as a Review in Advance on March 1, 2018

The *Annual Review of Clinical Psychology* is online at clinpsy.annualreviews.org

https://doi.org/10.1146/annurev-clinpsy-050817-084746

Keywords

treatment selection, precision medicine, personalized medicine, stratified medicine, depression, mental health treatment

Abstract

Mental health researchers and clinicians have long sought answers to the question "What works for whom?" The goal of precision medicine is to provide evidence-based answers to this question. Treatment selection in depression aims to help each individual receive the treatment, among the available options, that is most likely to lead to a positive outcome for them. Although patient variables that are predictive of response to treatment have been identified, this knowledge has not yet translated into real-world treatment recommendations. The Personalized Advantage Index (PAI) and related approaches combine information obtained prior to the initiation of treatment into multivariable prediction models that can generate individualized predictions to help clinicians and patients select the right treatment. With increasing availability of advanced statistical modeling approaches, as well as novel predictive variables and big data, treatment selection models promise to contribute to improved outcomes in depression.

Contents

INTRODUCTION

Depression is the world's leading cause of disability (World Health Organization 2017). Despite the existence of a variety of evidence-based interventions for major depressive disorder (MDD), response rates in the treatment of depression remain approximately 50% (National Health Service 2016, Papakostas & Fava 2010). The pursuit of novel neurological (e.g., deep brain stimulation; Mayberg et al. 2005), pharmacological (e.g., ketamine; McGirr et al. 2015), and psychological (e.g., positive affect treatment; Craske et al. 2016) treatments is one avenue through which researchers are attempting to improve treatment outcomes (Holmes et al. 2014). This review focuses on an alternative approach: treatment selection, the aim of which is to provide for each individual the treatment, among the available options, that is most likely to lead to a positive outcome for them.

Half a century ago, Gordon Paul (1967) stated, in a paper that has been cited more than 1,000 times: "[i]n all its complexity, the question towards which all outcome research should ultimately be directed is the following: What treatment, by whom, is most effective for this individual with that specific problem, and under which set of circumstances?" The spirit of this passage—the question "What works for whom?"—has been invoked in countless discussions of evidence-based practices in clinical psychology.

The idea is a good one, recognizing that no single treatment is likely to be the best for everyone. How to address this issue, however, has not been obvious or simple. In recent years, researchers have developed and tested the utility of multivariable prediction models to address the "What works for whom?" question. The promise of this work lies in the ability of such models to integrate multiple sources of information, rather than to rely on a single feature to inform treatment selection. In other areas of medicine, the effort to match individuals to their indicated

treatments is called precision medicine (Hamburg & Collins 2010), which has largely replaced the term personalized medicine (Katsnelson 2013, Schleidgen et al. 2013). Precision medicine[1] has afforded major advances in cancer treatment (National Research Council 2011, Schwaederle et al. 2015). For example, chemotherapy is the standard treatment for non–small-cell lung carcinoma (NSCLC). Early trials of the drugs erlotinib and gefitinib found little to no benefit of these drugs alone or in combination with chemotherapy (Pao & Miller 2005). However, recent clinical trials have found significantly improved outcomes for these drugs, relative to chemotherapy, in a specific subset of NSCLC patients with tumor mutations linked to the mechanisms of action of erlotinib and gefitinib (Paez et al. 2004, Rosell et al. 2012). We believe that similar approaches can help improve outcomes in mental health.

In this review, we describe several approaches to selecting the right treatment for an individual with depression. A striking feature of efforts in this area is the heterogeneity of the statistical approaches that are employed (Petkova et al. 2017, Weisz et al. 2015).

Variables that predict outcome are of two kinds: prescriptive or prognostic. Prescriptive variables, often referred to as moderators, affect the direction or strength of the differences in outcome between two or more treatments (Baron & Kenny 1986), and thus can help predict whether a patient will benefit more from one treatment relative to another. Cronbach (1957) described prescriptive relationships as "aptitude-by-treatment" interactions, which have typically been explored through subgroup or subset analysis (Doove et al. 2014, Wang & Ware 2013). **Figure 1** displays a variety of types of prescriptive relationships, which can be ordinal (sometimes called quantitative interactions) or disordinal (sometimes called qualitative interactions; involving a full crossover) (Gail & Simon 1985, Gunter et al. 2011a, Wellek 1997, Widaman et al. 2012). Fournier et al. (2008) reported an example of a disordinal moderator in depression: The presence of a comorbid personality disorder (PD) predicted better response with antidepressant medication (ADM), relative to cognitive therapy (CT), and its absence predicted a better response to CT than to ADM.

A variable is prognostic if it predicts response in a single treatment, or irrespective of treatment condition. If only one intervention is being analyzed, only prognostic relationships can be inferred. Although a predictor[2] may appear to be prognostic in a single-treatment analysis, it might predict differential treatment response in a study that compares two or more treatments. Additionally, a variable can function as a prognostic predictor in one context and as a prescriptive predictor in another. For example, higher depression severity is associated with worse outcomes in depression. In comparisons of medication to CT, baseline severity is prognostic because it has the same relationship to outcome in both treatments (Weitz et al. 2015). However, in comparisons of medication to placebo (Ashar et al. 2017) or of psychotherapies to control conditions, higher baseline severity predicts a larger advantage of the active treatment over the control, making severity prescriptive in these contexts (Driessen et al. 2010, Fournier et al. 2010).

[1]As defined by a National Research Council report, precision medicine "refers to the tailoring of medical treatment to the individual characteristics of each patient. It does not literally mean the creation of drugs or medical devices that are unique to a patient, but rather the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease, in the biology and/or prognosis of those diseases they may develop, or in their response to a specific treatment. Preventive or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not. Although the term 'personalized medicine' is also used to convey this meaning, that term is sometimes misinterpreted as implying that unique treatments can be designed for each individual" (National Research Council 2011, p. 125).

[2]The term predictor is sometimes used to refer specifically to prognostic relationships, but it can also be used to refer broadly to both prognostic and prescriptive variables, which is the way we use it here.

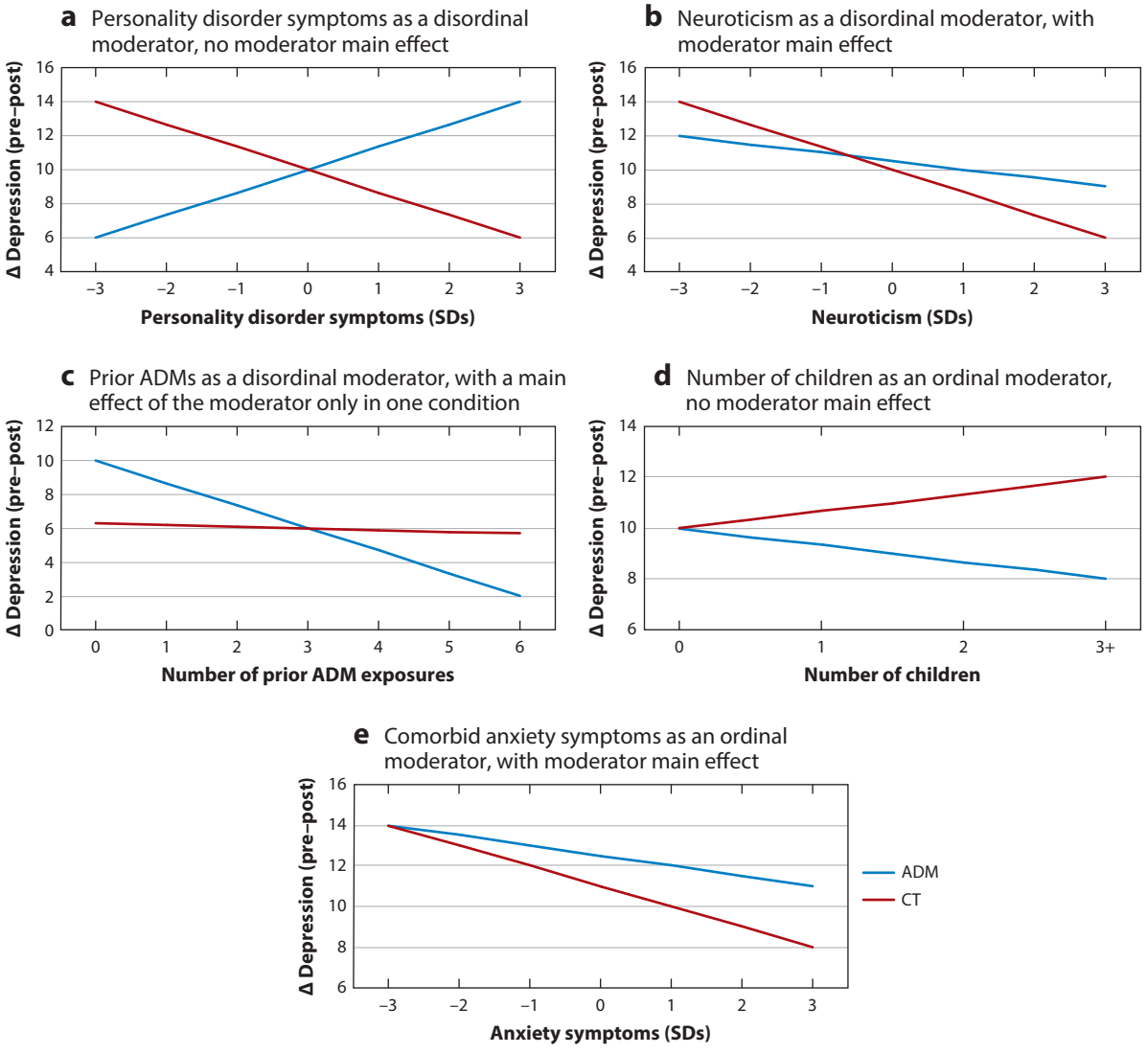

Figure 1

Five ways in which the differential effects of two treatments can vary as a function of a continuous moderator variable, and in which the interactions between treatment and moderator are linear. The relationships shown are for illustrative purposes only, but they draw on observations in the relevant empirical literatures. In *a–c*, at high levels of the moderator, one treatment is expected to produce stronger effects, whereas at low levels the other treatment is expected to be superior (disordinal interactions). In *d* and *e*, one of the two treatments is superior, on average, but the degree of superiority is expected to vary with the level of the moderator (ordinal interactions). In *b*, *c*, and *e*, the moderator is also a prognostic variable, such that a score on the moderator predicts outcome, independent of treatment (moderator main effects). In *a* and *d*, there is no moderator main effect. The moderator is predictive only in concert with treatment. Higher change scores on the y-axis indicate more improvement. Abbreviations: ADM, antidepressant medication; CT, cognitive therapy; SD, standard deviation.

Therapists select treatments for patients as a matter of course in every day practice. A clinician who attends to information about a specific client's presentation will likely generate hypotheses about the client's expected responses to potentially available treatments (Lorenzo-Luaces et al. 2015). These predictions may draw on a variety of sources, including a clinician's history with clients with similar features, their experiences in training and supervision, reasoning based on theory, and the empirical literature on treatment response (Raza & Holohan 2015). However, there is limited empirical literature to guide personalized, or precision, selection of treatments. Clinicians are therefore forced to practice what Perlis (2016) has dubbed "artisanal medicine."

Artisanal medicine is the practice of making treatment decisions in an idiosyncratic or unsystematic manner, or in a manner guided by theory and experience but largely uninformed by empirical evidence or feedback. Unfortunately, the lack of standardization that defines artisanal medicine limits the validity and utility of such approaches for decision making (Dawes 1979, 2005; Dawes et al. 1989; Tversky & Kahneman 1983).

In treatment contexts, statistical decision making, also called actuarial decision making, relies on predictions made with the use of algorithms, in a reproducible way. Grove et al. (2000) detail the ways in which actuarial approaches to decision making can overcome limitations and biases prevalent in human judgment (Dawes et al. 1989, Pauker & Kassirer 1980). By and large, empirical tests of clinical versus actuarial prediction (Grove & Meehl 1996) have revealed the superiority of actuarial methods. More than 60 years ago, Paul Meehl (1954) published his seminal monograph on this topic, titled *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. The field of mental health treatment has only just begun to apply Meehl's line of thinking to precision mental health.

For much of the twentieth century, evidence-based practice in mental health has largely concerned the provision of a specific treatment to patients based on a specific DSM-defined disorder. Evidence to guide such decisions has come from randomized clinical trials (RCTs), in which active treatments are compared to control conditions or to other active treatments (Chambless & Hollon 1998). For example, on the basis of positive findings in RCTs, CT (Beck et al. 1979) and interpersonal therapy (IPT; Klerman & Weissman 1994) are each considered evidence-based psychotherapies for MDD. Similarly, among the psychoactive medications, specific classes of drugs have been studied under the assumption that they have differential efficacy with specific disorders (Fineberg et al. 2012).

However, as has been widely discussed in the literature, the library of empirically supported treatments (ESTs) is insufficient to address clinician and client needs (Hollon et al. 2002). The average treatment effect (ATE) is the extent to which, for the clients in the sample, a given intervention leads to more (or less) symptom improvement, on average, relative to comparator conditions. The main findings from RCTs refer to effects of treatments, on average, and not to potentially important sources of variability in treatment response (Imai & Ratkovic 2013, Kessler et al. 2017). Consider, for example, a case in which an ATE of 10 points of change in the Beck Depression Inventory (BDI) is estimated in a comparison of a strong treatment versus a weaker intervention. This might reflect an average change of 20 in the strong condition and 10 in the weak condition. It would be a mistake to assume that, even with such a large average change in the strong treatment condition, every client benefited substantially more from it than they would have from the weak condition. In fact, it is typical in such studies that reductions in the BDI observed in some clients are near the group average, whereas in others the reductions will be quite large, and in still others there will be little or no change observed. Understanding the heterogeneity of treatment effects could facilitate treatment selection by identifying individuals for whom more than a 10-point advantage of the stronger treatment would be expected, as well as individuals for whom the weaker treatment might be equally, or even more, effective (Kessler et al. 2017).

Supplemental Material

INTRODUCTION TO RESEARCH ON PREDICTIVE VARIABLES

Overview

Variables examined in research on the prediction of mental health outcomes in depression have been drawn from a variety of sources, including routinely assessed domains such as demographic, environmental, and diagnostic information. A recent emphasis on neurobiological variables (Gabrieli et al. 2015, Jollans & Whelan 2016, Leuchter et al. 2009, Pizzagalli 2011, Stephan et al. 2017) has begun to reveal the potential of the inclusion of neurocognitive (Gordon et al. 2015) and biomarker-based assessments as predictors (Uher et al. 2014). For example, McGrath et al. (2013) measured pretreatment brain activity using positron emission tomography in a depression RCT and identified right-anterior-insula metabolism as a disordinal moderator: Insula-hypometabolism was associated with better outcomes in CT and worse outcomes with ADM, while insula-hypermetabolism was associated with the opposite pattern. In **Supplemental Table 1** we list recent reviews of predictors in depression, the results of which could inform future treatment selection investigations.

By far, the most common approach to the prediction of treatment response in mental health is to take advantage of the information captured in prognostic relationships (e.g., Rubenstein et al. 2007). Prognostic statements regarding response to intervention are of the following form: A client with characteristic X, in a given context (e.g., with any intervention, or with a specific treatment[3]), has a Y% chance of experiencing symptom remission. Prognostic information can be used to provide realistic expectations to the treating clinician, as well as the client and their family (Kessler et al. 2016). This includes expectations concerning the rapidity and extent of response to the treatment that will be provided, as well as whether special attention should be paid to the client's progress (Hunter et al. 2011, Lutz et al. 2014).

The information conveyed in a prognostic statement does not inform directly the following question: "What is the best available option for this client at this time?" A common mistake in the interpretation of a prognostic finding is to conclude that clients found to have a poor prognosis in a given treatment will have a better prognosis in a different treatment (Simon & Perlis 2010). Consider the finding that, in CT, patients with chronic depression exhibit lower recovery rates than those with nonchronic depression (Fournier et al. 2009). This might indicate that other interventions (e.g., ADMs) or treatments created specifically for chronic depression, such as Cognitive Behavioral Analysis System of Psychotherapy (CBASP; McCullough 2003), should be preferred over CT for individuals with chronic depression. However, it could instead be that CT is as effective as (or even more effective than) other available interventions for such individuals (Cuijpers et al. 2017). Indeed, evidence from an RCT comparing CT to ADM suggested that chronicity is prognostic, in that it was associated with similarly lower response rates in both treatments (Fournier et al. 2009), and an RCT comparing CBASP to ADM in individuals with chronic depression found no difference in response rates (Nemeroff et al. 2003). The only type of investigation that can directly address the prescriptive question (i.e., "Which treatment is likely to be most effective for a client with X, Y, and Z characteristics?") is one that focuses on moderation.[4] Unfortunately, analyses of this type are much less frequently conducted (see

[3]We are referring to a case in which a characteristic predicts outcome in studies of a single treatment, and in which its predictive value is unknown in other contexts. Note that if a factor predicts outcome in one treatment but does not predict outcome in a second, it could be prescriptive in that context (see **Figure 1*c*** for an example).

[4]Prescriptive questions can be investigated through the simultaneous use of two or more prognostic models in the same sample (e.g., see Kessler et al. 2017).

Supplemental Example 1 for an early example of the single moderator approach, with a twist, from Beutler et al. 1991).

Supplemental Material

Studies in which pretreatment variables are found to predict treatment response can provide clues about treatment mechanisms (typically identified in efforts to find variables that mediate a treatment effect) (MacKinnon et al. 2007), and thus can help distinguish between compensation and capitalization models of the effects of psychotherapies. The compensation model is that individuals with deficits in areas targeted by a therapy will benefit the most from it. An example of this is the hypothesis that CT, which targets dysfunctional cognitions, would be preferred over IPT for individuals high on cognitive dysfunction and low on problems of interpersonal functioning, and vice versa. The support for this hypothesis is equivocal. Capitalization models, which propose that therapies work best when they build on clients' strengths, have received some support (Barber & Muenz 1996, Cheavens et al. 2012).

Understanding Moderator Relationships

Given the observed heterogeneity in the presentation, history, and prognosis of depression, it is unlikely that any single variable in isolation will have clinically useful predictive utility (Simon & Perlis 2010). Nonetheless, considering how a single moderator would guide treatment selection is a useful exercise for enhancing one's understanding of how multivariable treatment selection algorithms work. To that end, we created plots (see **Figure 1**) depicting hypothetical examples[5] of prescriptive relationships that could be observed with continuous moderators. In **Supplemental Figure 1**, we also discuss the application to clinical decision making of findings of prescriptive relationships when the moderators are binary.

In empirical reports of moderator findings, the distinctions between different types of prescriptive relationships illustrated in **Figure 1** are rarely made, and when the details of these relations are implied they can be inconsistent, misleading, or incorrect. Issues with data processing and the behavior of regression coefficients can make interpreting and describing moderator relationships difficult even for the individuals who perform the analyses (Kraemer & Blasey 2004). To learn more about these topics, we refer the reader to Kuhn & Johnson (2013).

Consider the following statement: "Clients high on characteristic Z experienced superior outcomes with ADM, relative to CT." It is tempting to infer that those who are low on characteristic Z would respond better to CT than to ADM, but there is nothing in the statement about such individuals. Thus, the statement could describe any of a variety of relationships, including those that **Figure 1*a*,*b*,*e*** depicts. If **Figure 1*e*** were true, a clinician should encourage individuals with high levels of Z to pursue ADM, whereas individuals with low levels of Z should be informed that there is no indication of a meaningful difference between the two treatments. However, the relationship could also be characterized by the pattern in **Figure 1*a*,*b***. If this were true, individuals high on Z would receive the same recommendation (choose ADM), but individuals low on Z should be steered away from ADM and toward CT. In the case of **Figure 1*a***, an individual with an average level of Z would be informed that the two treatments are expected to be similarly effective for him or her, and the expected size of the advantage of one treatment over the other is similar at each end of the spectrum. In the case of **Figure 1*b***, the expected advantage of ADM over CBT for those high on Z is larger than the expected advantage of CBT over ADM for those

[5]These examples are for illustrative purposes only. We drew upon patterns that have been observed in empirical studies, but the figures do not represent empirical findings, per se. We followed the structure used by Kraemer (2013) and Schneider et al. (2015) in creating these figures.

low on Z. This example illustrates one of the many ways in which the translation from analysis to interpretation to implementation can result in either optimal, suboptimal, or even harmful application of prescriptive information to clinical decision making.

The importance of evaluating the evidence for predictors prior to utilizing them in clinical settings deserves special emphasis in treatment selection (Howland 2014, Perlis 2016). When reading an empirical investigation of individual differences in treatment response, one must identify the population from which the sample was drawn. Although a paper may describe its findings as pertaining to "treatment response in depression," it is necessary to attend to specific features of the sample (e.g., inclusion/exclusion criteria, range of depression severity, extent of comorbidity, treatment history) to determine the pertinence of the evidence to a specific client. For example, depressive symptom severity has been reported to predict differential response to ADMs versus placebos, with ADMs evidencing superiority over placebo for moderate to high severity, and little to no differences seen at the lower end of the severity spectrum (Barbui et al. 2011, Fournier et al. 2010, Khan et al. 2002, Kirsch et al. 2008). However, for most trials, entry criteria include moderate or greater symptom severity (Zimmerman et al. 2015, 2016), thus restricting the range of severity that can be investigated, and constraining the applicability of many positive moderator findings to a subset of the population of patients with MDD. There are many examples of predictive algorithms built using data from a sample of clients treated with one antidepressant that have failed to generalize to a different antidepressant (Chekroud et al. 2016, Iniesta et al. 2016a, Perlis et al. 2010). Similarly, models predicting the onset of major depressive episodes in European primary care (King et al. 2008) have not generalized to the US general population (Nigatu et al. 2016).

Reliance on tests of significance can result in misleading impressions about the importance of predictive variables (Nuzzo 2014, Wasserstein & Lazar 2016). For example, if a variable selection approach relies on p-values (often with $p < 0.05$ as a threshold) to assess statistical significance, a variable could miss a predetermined cut-off by a small margin, leading to a report that the variable is not predictive (Bursac et al. 2008). However, the difference in the predictive utility of an excluded variable that "just missed" (e.g., $p = 0.06$) and an identified predictor that is "barely" significant (e.g., $p = 0.04$) is, of course, trivial (Mickey & Greenland 1989). Most RCTs are powered to detect main effects, and therefore are powered to detect only very strong interactions. Complicating matters further, different analytic approaches can identify different variables, even when applied to the same data (Cohen et al. 2017). Additionally, variables that were not assessed, or that were assessed and not analyzed, could also be important predictors. Finally, statistically significant results are not necessarily clinically significant, if effect sizes are small (Meehl 1978). In the context of a large sample, small or weak relationships can be identified as statistically significant. However, statistically significant variables are not always good predictors (Lo et al. 2015). More relevant are metrics that can characterize the importance of the relationship and can therefore quantify and translate the clinical meaning of the findings (Bossuyt & Parvin 2015). For example, Janes et al. (2011) developed a statistical method for evaluating treatment selection markers that went beyond the classic approach of testing for a statistical interaction between a predictor and treatment to answer four important questions: "1) Does the marker help patients choose among treatment options?; 2) How should treatment decisions be made that are based on a continuous marker measurement?; 3) What is the impact on the population of using the marker to select treatment?; and 4) What proportion of patients will have different treatment recommendations following marker measurement?" (p. 253). Moving beyond statistics, consideration of factors such as cost, feasibility, and client burden should be weighed against the additive predictive power of variables that exceed those routinely collected in clinical settings (Perlis et al. 2009).

Two Frequently Cited Treatment Selection Variables

In real-world contexts, two variables often influence treatment selection in depression. The first is client preference (McHugh et al. 2013): Many treatment guidelines (Hollon et al. 2014) specify the importance of attending to clients' preferences. However, studies of the predictive utility of client preference include positive (Kocsis et al. 2009, Mergl et al. 2011, Swift & Callahan 2009, Swift et al. 2011), mixed (Dunlop et al. 2017, Preference Collaborative Review Group 2008, McHugh et al. 2013), and negative (Dunlop et al. 2012b, Leykin et al. 2007b, Renjilian et al. 2001, Winter & Barber 2013) findings. Seemingly, contrary to lay intuition, preference is not a reliable indicator of treatment response. What's more, patients' preferences might shift when given individualized information about expected outcomes.

Second, a client's experience with previous treatments for depression can serve a prognostic or prescriptive function, as suggested by findings from several outcome studies. Prior exposure to ADMs and history of nonresponse to ADMs have each been found consistently to predict poor response to future courses of antidepressants (Amsterdam et al. 2009, 2016; Amsterdam & Shults 2009; Byrne & Rothschild 1998). Moreover, there is evidence that the number of prior ADM exposures can provide prescriptive information. For example, Leykin et al. (2007a) found that multiple previous ADM exposures predicted a poorer response to ADM, but not to CT, such that a client with two or more prior exposures was significantly more likely to benefit from CT than ADM. Clearly, assessing treatment history is important and could be used to inform treatment selection.

THE PERSONALIZED ADVANTAGE INDEX APPROACH

In 2011, we, along with other members of our research team, began to explore the possibility that machine learning[6] (Iniesta et al. 2016b, Passos et al. 2016) or multivariable regression modeling approaches could be brought to bear on problems in precision mental health. We initiated our journey with a specific goal in mind: to find or develop an approach that could identify clients with MDD for whom antidepressants are likely to be more beneficial than CT, and vice versa. Two findings from our lab prompted our interest. First, in a sample of clients with moderate to severe MDD, ADM and CT had produced nearly identical group-average effects on depressive symptoms over the course of a 16-week RCT (DeRubeis et al. 2005). Second, we had discovered five variables (marital status, employment status, PD comorbidity, antidepressant treatment history, and number of recent stressful life events) that served as moderators of symptom change in this sample (Fournier et al. 2009).

What was striking about the variables that moderated the effects of ADM versus CT was that no single one dominated the differential predictions. To survive the variable selection procedure, each variable had to make an independent contribution to the statistical model. As a result, the variables needed to be relatively uncorrelated with each other in the sample, such that they could not be used to define a factor, per se. Rather, we had identified five vectors, represented by the five variables, any one of which could be used to point a client to either ADM or CT as his or her preferred treatment, although there was not an especially strong predictor in the bunch. We understood this to indicate that there are many "reasons" one treatment may be more effective than another for a given person.

This posed a challenge for selecting treatments for patients with contradicting indications. For example, as noted above, clients with comorbid PD improved more with ADM than they did in

[6]Gillan & Whelan (2017) explain the following: "Machine-learning (essentially synonymous with 'data-mining' or 'statistical learning') refers to a class of approaches that focus on prediction rather than interpretation or mechanism" (p. 35).

CT, whereas clients without comorbid PD improved more in CT than with ADM (Fournier et al. 2008). It was also the case that clients who were unemployed improved more in CT than with ADM. How is a clinician to use this information in recommendations to a client with comorbid PD (indicating ADM) who is unemployed (indicating CT)? How does the clinician integrate information when considering different recommendations from the other three variables when forming a treatment recommendation? The implication of the literature on actuarial versus clinical decision making, which has focused on prognosis, is that outputs from a well-constructed statistical method should be able to provide useful information in treatment selection contexts, as well.

We also reasoned that effective guidance for clinicians and clients would, ideally, be "graded," to reflect the likelihood that for some clients differential benefit would be expected to be quite substantial, for others it would be negligible, and for others in between. To address these challenges, we developed the Personalized Advantage Index (PAI) approach (DeRubeis et al. 2014a), which has since been featured in work both internal and external to our lab (Cohen et al. 2017, Huibers et al. 2015, Keefe et al. 2018, Vittengl et al. 2017, Zilcha-Mano et al. 2016; also see C. Webb, M. Trivedi, Z. Cohen, D. Dillon & J. Fournier, manuscript under review).

Essential to the PAI approach is the identification of variables in a dataset that predict differential response to two or more treatments. Once the variables have been identified, a multivariable statistical model that includes interaction terms representing the prescriptive variables[7] is constructed. A PAI for a given client is then calculated as the difference between his or her predicted outcomes in two treatments (treatment A and treatment B). To generate the prediction for treatment A, the client's values on the baseline variables, as well as the value representing treatment A, are inserted into the model. This is repeated, with the value of treatment B inserted into the model. The predicted value with treatment A in the model is compared with the predicted value under treatment B. The sign of the difference reflects the model-indicated treatment, and the magnitude of the difference reflects the magnitude, or strength, of the predicted difference. We return to a focus on the PAI approach more specifically in a discussion of issues of broader importance in treatment selection, following a review of literatures on a variety of prognostic and prescriptive multivariable approaches in mental health.

REVIEW OF THE LITERATURE ON MULTIVARIABLE PREDICTION MODELS

Overview

If a single predictive variable with a very large effect can be identified in a treatment context, application to practice is likely to be straightforward. In depression, however, such variables have not been identified consistently. In part for this reason, single variables have not found widespread use in treatment selection contexts. One exception to this is baseline symptom severity, which has been included in many practice guidelines as an indication that stronger treatments, or the combination of ADMs and psychotherapy, are to be preferred over lower-intensity interventions (American Psychiatric Association 2010, National Institute for Health and Clinical Excellence 2009). The status of baseline severity as a prescriptive variable has been supported primarily in comparisons of an active treatment with a control (Driessen et al. 2010, Fournier et al. 2010), but not in comparisons of two active treatments (Vittengl et al. 2016, Weitz et al. 2015).

[7] Some of the machine-learning models we have constructed do not include interaction terms per se, but they perform the same task of modeling differential response.

Multivariable models are more likely to yield powerful predictions (Perlis 2013), and they comport with our understanding of psychopathology and treatment response as complex, multiply determined phenomena (Drysdale et al. 2017). Unfortunately, the interpretation and application of multivariable models is less straightforward for the clinician than are single-variable approaches. To further complicate matters, it may be important not only to consider multiple variables simultaneously, but also to consider potential interactions among multiple variables (Tiemens et al. 2016). As new, more powerful modeling approaches become available (Kapelner & Bleich 2016, Luedtke & van der Laan 2016, Ma et al. 2016), researchers must weigh the increased flexibility and predictive power of such approaches against the interpretability (Hastie et al. 2009, James et al. 2013) of simpler models (Green & Armstrong 2015), especially insofar as the goal is to disseminate the models in ways that are acceptable to clinicians and clients (Delgadillo et al. 2016).

Prognostic Models

Recent multivariable modeling efforts (Chekroud et al. 2017) highlight the potential for these advanced approaches to improve prognostic prediction in mental health (see Gillan & Whelan 2017 for an extensive review). For example, Chekroud et al. (2016) used archival data from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study (Trivedi et al. 2006) to identify predictors of response to acute selective serotonin reuptake inhibitor (SSRI) treatment of depression. They validated their model in an external sample from a separate study, the Combining Medications to Enhance Depression Outcomes trial (Rush et al. 2011). Interestingly, although they found that it had acceptable predictive power in the study's two SSRI conditions, it was not significantly predictive for the validation sample's non-SSRI ADM condition, suggesting that the model may not work outside of the drug class on which it was developed.

It is understandable that more progress has been made to date with prognostic models, relative to prescriptive models, but the former have less relevance to a core question in mental health treatment: "Which treatment should client X pursue to have the greatest likelihood of having a positive outcome?" In many medical contexts, treatment recommendations follow from accurate diagnosis. However, for many mental health diagnoses, especially depression, there exists an abundance of ESTs that a client could potentially receive. Thus, in mental health, using person characteristics to help guide individuals to their optimal treatment is especially important. Uher et al. (2012) noted the limitation of their prognostic model, which predicts response to ADM in MDD, relative to a prescriptive model: "Clinical application of this finding will require identification of a treatment that is effective in individuals [predicted to have poor response to ADM]" (p. 976).

TREATMENT SELECTION APPROACHES

Overview

Byar & Corle (1977) published an early example of a multivariable treatment selection model in medicine. Working with longitudinal data from a sample of men who were randomized to one of two treatments for prostate cancer, they explored whether, for each man, the more promising treatment of the two could be identified using a set of characteristics ascertained prior to random assignment. At the time, the field's emphasis had been on discovering, for all patients with a given diagnosis, "which treatment is best." Byar & Corle capitalized on advances in statistical methodology that allowed for survival modeling with multiple covariates and used the heterogeneity of patients to develop a rubric that could, in principle, inform individual treatment recommendations. Byar (1985) later applied this general approach to the differential prediction of survival in response to two dosage levels of chemotherapy for prostate cancer. Surprisingly, not until 1994 was any of

Byar's work cited by others in the context of actuarial modeling in treatment selection. Yakovlev et al. (1994) applied a similar methodology to a treatment selection problem in cervical cancer, but from 1994 until 2011 (Gunter et al. 2011b) none of these works was cited in a publication that applied or extended the differential prediction methods described by Byar or Yakovlev.

The past half-decade has witnessed a surge of interest in optimizing treatment selection using multivariable predictive models, and much of this work is focused on treatments for depression. When moving beyond prognostic prediction into treatment selection, several additional considerations come into play. The first factor to consider is whether the treatment decision is between two or more equivalent interventions or, instead, between a stronger versus a weaker intervention, as this distinction has implications for how one builds and evaluates the models. We begin our discussion focusing on contexts in which the decision is between equally effective treatments, when the question truly is "What will work best for each given patient?" We follow this with a review of the special case of stratified medicine (in the context of stepped-care), in which at least one of the candidate interventions results in greater improvement than a comparison condition, on average.

One of the earliest examples of multivariable treatment selection in mental health came from Barber & Muenz's (1996) reanalysis of the Treatment of Depression Collaborative Research Program (Elkin et al. 1989) study, which compared CT to IPT for MDD. The authors built a "matching factor" that combined the prescriptive value of marital status, avoidance, obsessiveness, and baseline severity in a linear model predicting symptom change. They also tested the prescriptive value of two personality disorder diagnoses, avoidant PD and obsessive-compulsive PD, and proposed that the models including these factors could be used to match patients to CT or IPT.

Lutz et al. (2006) employed a statistical technique called "nearest-neighbors" to predict differential outcomes between two variations of CT. In the nearest-neighbors method, each client's outcome in each treatment is predicted from the average observed outcomes in the respective treatments of groups of clients who are most similar to the index client on a set of features.

Kraemer (2013) proposed a method that involves the creation of a single variable (termed M^*) that represents a weighted combination of multiple moderators. This approach was demonstrated using data from a randomized comparison of IPT versus ADM (Wallace et al. 2013). The statistical approach behind the M^* method excludes any consideration of main effects in an attempt to maximize the power of the differential prediction of outcome (Kraemer 2013). Thus, two clients with identical M^* scores could have very different prognoses, but this information is not given by the method. Recently, the M^* approach has been used to analyze data from a comparison between aripiprazole augmentation and placebo augmentation for ADM-treatment-resistant late-life depression (Smagula et al. 2016), and between two psychological treatments for clients with anxiety disorders (Niles et al. 2017a,b).

A series of papers by Uher et al. demonstrates the evolution of treatment selection from single to multivariable approaches. Using data from the Genome-Based Therapeutic Drugs for Depression study (Uher et al. 2009), they tested the prognostic and prescriptive utility of three symptom clusters (factors) and the six symptom dimensions that made up the factors (Uher et al. 2012). They examined the predictive power of each of these nine variables in isolation and found evidence for only the anxiety symptom dimension as a moderator. Recently, they returned to the question of treatment selection in this sample, using a multivariable approach with an expanded set of potential variables (Iniesta et al. 2016a). They found that a model that simultaneously included the effects of multiple variables could predict differential response to antidepressants with clinically meaningful accuracy, thus demonstrating the potential of multivariable approaches for treatment selection.

Other groups have used variants of the methods already described to address treatment selection questions (Cloitre et al. 2016, Westover et al. 2015). In **Supplemental Table 2** we contrast some of the approaches used in the multivariable prediction work referenced in this review. Although this

Supplemental Material

abundance demonstrates the strong interest in precision medicine, the heterogeneity of methods (Doove et al. 2014) contributes to difficulties in detecting consistencies and inconsistencies in predictors, and creates a barrier to identifying "best practices."

To date, most attempts to build prescriptive models for treatment selection have utilized data from RCTs. Future efforts, exemplified by the ongoing work of Gillan et al. to collect mental health treatment outcome data online (Gillan & Daw 2016), will likely also rely on nonrandomized data. The potential influence of unknown confounds is a limitation of treatment selection efforts outside the context of RCT data. The bias in predictions in such studies can derive from "selection effects," which result when clients with a given feature (e.g., history of nonresponse to ADMs) are provided with one of the treatments preferentially (e.g., CT). In these contexts, approaches such as propensity score analysis (d'Agostino 1998) can be employed to mitigate the effects of confounds.

Extending the Personalized Advantage Index to Stratified Medicine

When a model is developed to guide treatment decisions in health care contexts in which the available interventions differ in terms of strength, cost, availability, or risk, the question "Which treatment is predicted to be most effective for each individual?" may be moot. The treatment with the strongest effect on average is likely to be predicted to be the most effective one for most or all of the clients. In these contexts, the more relevant question is often "What is the best way to allocate the stronger/costlier/less available/riskier (hereafter 'stronger') treatment?" The practical goals of predictive models in stratified medicine are to enhance the efficient allocation of scarce or costly resources, as well as to limit patients' unnecessary exposure to treatments that require substantial time commitments or are associated with heightened side effect risk (Hingorani et al. 2013).

Considerations of treatment allocation for stronger versus weaker interventions, including part-whole comparisons (e.g., combined ADM + CT versus ADM alone), should address the distinction between two ways in which the stronger treatment produces superior average change. One possibility is that every client is expected to benefit more—and by a similar amount—from the stronger treatment. In such cases, decisions about who should be provided the stronger treatment will not be based on clients' predictive features, except insofar as the clients with the worst prognoses might be provided the strongest treatment, for ethical reasons. However, it may be that individuals vary in regard to how much more they will benefit from the stronger treatment, relative to the weaker one. In such cases, it becomes important to identify client characteristics that predict differential response to the stronger versus the weaker treatment.

Patient Subtypes

To better describe the patient types on whom treatment selection might be tested, we propose an adaptation of DeRubeis et al.'s (2014b) conceptualization of client types. In an attempt to highlight the relationship between therapy quality and patient improvement, they posited five exemplar types meant to represent the spectrum of potential associations that could be expected between therapy quality and improvement: spontaneous remitters, easy patients, pliant patients, challenging patients, and intractable patients. For spontaneous remitters, any level of therapy quality (from the best to the worst) would lead to high levels of improvement. For patients at the other end of the spectrum (the intractable patients), little to no improvement would be expected, regardless of the level of therapy quality. In the middle of this spectrum are pliant patients, defined as those patients whose improvement would vary as a function of therapy quality, such that with very poor quality therapy or no therapy, no improvement would be expected, and with the highest quality therapy possible, complete improvement would be expected to result. For the purpose of treatment selection, the pliant patient category can be broken down into two subgroups: individuals

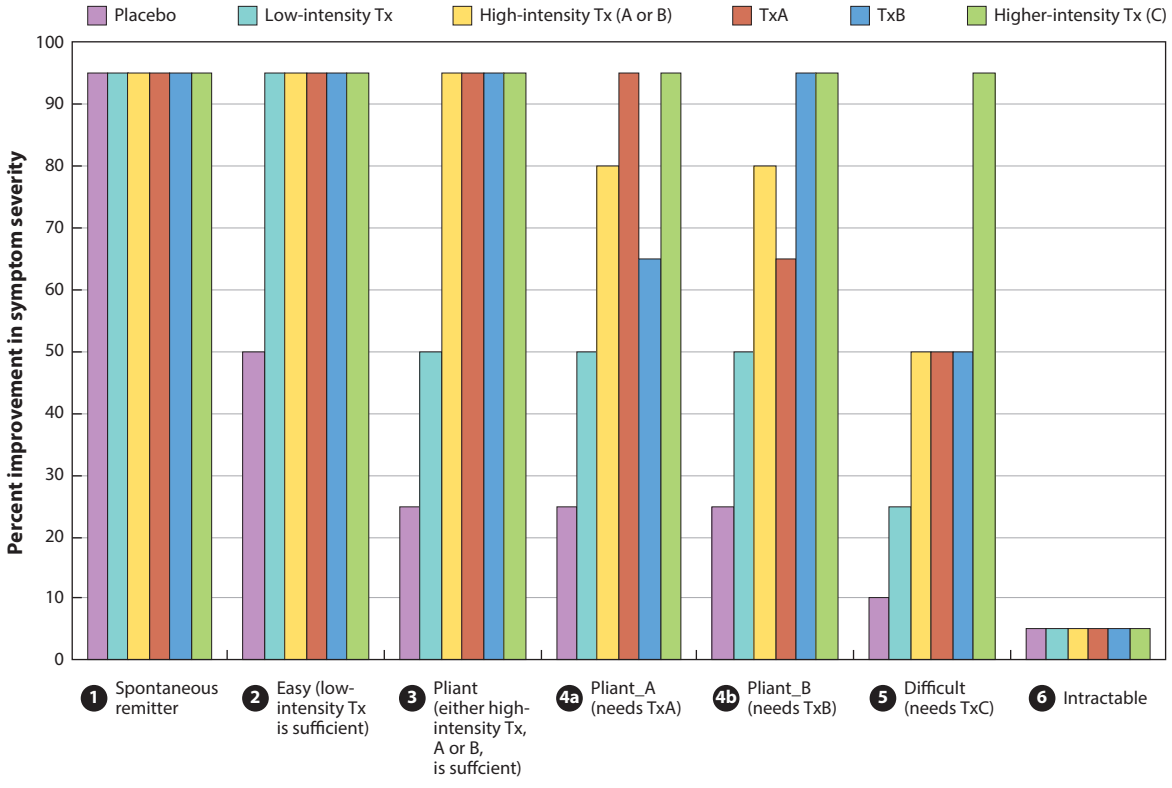

Figure 2

The figure depicts the expected improvement for different patient prototypes in different treatment contexts. The treatment contexts range from lowest to highest intensity (colored bars). Patient prototypes, which range from spontaneous remitters to intractable patients, are labelled on the x-axis. As shown with the colored bars, spontaneous remitters would be expected to show the same high level of response (95%) in any treatment context. Similarly, intractable patients would be expected to show the same low level of response (5%) irrespective of the treatment provided to them. Prototypes 2, 3, 4a, 4b, and 5 would be expected to show different levels of response depending on the treatment provided. Prototypes 3, 4a, and 4b are all "pliant," but they differ in regard to the expected responses to the two high-intensity treatments (TxA and TxB). Patients represented by prototypes 4a and 4b differ from those represented by prototype 3 in that they require a *specific* high-intensity treatment, whereas prototype 3 patients would be expected to evidence a high level of response to either high-intensity treatment. This distinction is also depicted by the heights of the yellow bars (unspecified high-intensity treatment), which represent the averages of the expected responses to TxA and TxB within each prototype.

who would improve insofar as they receive quality treatment of any type and other individuals for whom the correlation between quality and improvement would be moderated by the "match" between patient and treatment (see **Figure 2**, types 3 and 4a/4b). These latter individuals (**Figure 2**, type 4a/4b), who will respond well to—but only to—a specific treatment, are the individuals for whom PAI-type treatment selection will be most important.

The analytical tools used to construct PAI models can be adapted to inform decisions in stratified medicine, where the choice is often between a high- versus low-intensity treatment, and where the high-intensity treatment is more effective, on average. In such cases, the goal is to distinguish between individuals who are likely to benefit much more from the high-intensity treatment than from the low-intensity treatment, versus those for whom the expected differential benefit is small. As with the PAI approach, a continuous index is created (Forand et al. 2017), but in this case its purpose is to array patients along a continuum from those who are most likely to

experience a positive response irrespective of treatment to those for whom the expected outcome is poor (**Figure 2**, types 2/3/5). Lorenzo-Luaces et al. (2017) implemented such an approach as a proof of concept, with data from a randomized comparison of a high-intensity treatment (CT) with two lower-intensity treatments. On average, as described in the main outcome paper from the trial, the differences between CT and each of the two comparison conditions were small (van Straten et al. 2006). Lorenzo-Luaces et al. constructed a multivariable prognostic index[8] as described above. Following DeRubeis et al. (2014b), they predicted that, for clients with poorer prognoses, the provision of CT would lead to a higher likelihood of response, relative to the lower-intensity conditions. Between-treatment comparisons were not expected to reveal differences in response rates in the subset of clients with scores indicating better prognoses. Findings were consistent with these predictions, suggesting that the application of these principles in stratified medicine could substantially increase the efficiency of mental health treatment systems. Gunn et al.'s (2017) recently initiated RCT tests a symptom-based depression clinical prediction tool called Target-D for stepped-care in primary care.

Two recently published works using data from the National Health Service's Improving Access to Psychological Therapy (IAPT) program also highlight ways in which multivariable models could be used to guide stratified medicine in mental health. Saunders et al. (2016) used latent profile analysis to create eight profiles that described sets of baseline demographic data and symptom features that defined patient clusters. One of their goals was to identify subsets of clients (those with profiles similar to each other) for whom differential predictions could be made between high-intensity treatment and low-intensity psychological treatment. In a different sample of clients treated for mood and anxiety disorders in the IAPT services, Delgadillo et al. (2016) explored the potential utility of treatment selection models. The authors created an index that generated predictions as to which clients were likely to achieve reliable and clinically significant improvement in depression or anxiety symptoms. Recent work using a prognostic index of case complexity yielded similar results in a separate sample of IAPT patients (Delgadillo et al. 2017).

RECOMMENDATIONS FOR BUILDING TREATMENT SELECTION MODELS

In what follows, we review the major steps involved in constructing and evaluating a treatment selection approach from a dataset that includes values, for each client, on variables that reflect pretreatment characteristics, the treatment provided to the client, and the client's observed outcome in that treatment. Understanding these steps is critical for the clinical researcher who wants to conduct PAI analyses, as well as for the clinician who wants to interpret and evaluate the utility of findings from treatment selection studies.

The first step is to identify and prepare the candidate predictor variables. Good candidate predictor variables are those that are measured prior to the point of treatment assignment and that plausibly could be related to outcome, either in general (prognostic) or differentially between treatments (prescriptive). If prior research has indicated that a variable predicts outcome, then it should be included as a potential predictor, but as the literature on predictors (and especially on moderators) in mental health is still relatively sparse, including other putative variables is recommended. Variables must not have significant missingness, and tests for systematic missingness

[8] A prescriptive index could also be used in the context of such a comparison. Use of a prescriptive model in this context would identify patients for whom the stronger treatment is expected to lead to better outcomes than the weaker treatment, patients for whom less advantage of the stronger treatment is expected, and perhaps a subset of patients for whom no advantage of the stronger treatment is expected, or even a subset for whom the weaker treatment is predicted to be better than the stronger treatment.

should be performed to inform the appropriateness of imputation (Jamshidian & Jalal 2010). Variables should also exhibit sufficient variability. For example, it makes little sense to include gender if 95% of the sample were female. Many variable selection and modeling techniques used in prediction are sensitive to situations in which the set of predictors has high collinearity, and thus it is wise to examine the covariance structure of the potential predictors, and to take steps to reduce high collinearity (Kraemer 2013). Other considerations for preparing potential predictors include dealing with outliers/leverage points, making categorical variables binary (where indicated), and transforming variables for theoretical reasons or to deal with problematic distributions (e.g., those with high skew). Finally, centering variables can help avoid inferential errors and increase stability when using regression-based approaches (Kraemer & Blasey 2004).

The choice of variable selection and modeling approaches can be constrained by the nature of the outcome variable. Although many approaches can accommodate both binary and continuous outcomes, the use of categorical outcomes, or longitudinal and survival-type outcomes, is limited to a select subset of the available approaches.

Once potential predictor and outcome variables have been selected, the next step is to build the prediction model. This is typically a two-step process comprising variable selection and model-weight specification. Many different variable selection approaches have been proposed for treatment selection, all of which attempt to identify which variables, among the potential predictors, contribute meaningfully to the prediction of outcome. Gillan & Whelan (2017) provide an excellent discussion of theory-driven versus data-driven approaches to model specification. Classic approaches rely on parametric regression models [e.g., forward or backward stepwise regression; see Fournier et al. (2009) for a worked example] that select only those variables with statistically significant relations with outcome. A subset of these approaches includes penalties that aim to create parsimonious models by limiting the number of variables selected (Tibshirani 1996). Others employ bootstrapping procedures or shrinkage parameters that seek to maximize the stability and generalizability of the models (Austin & Tu 2004, Garge et al. 2013, Zou & Hastie 2005). Advances in statistical modeling and computational resources have led to feature selection approaches, many of which are based on machine learning, that can flexibly model and identify predictors with nonlinear and higher-order interactions (Bleich et al. 2014). The line between variable selection and model weight specification is not always clear, as some modeling approaches combine the two in one step. Gillan & Whelan (2017) provide an in-depth review of the merits of machine learning in mental health; interested readers can also consult books focused on applied clinical predictive modeling (Chakraborty & Moodie 2013, Parmigiani 2002, Steyerberg 2008).

Cohen et al. (2017) proposed a new variable selection approach that combines the outputs of multiple procedures with the aim of generating robust predictors. It also allows for the inclusion of complex relations that often exist between predictors and outcome in treatment selection contexts that are often overlooked in classic regression-based approaches. We performed four different variable selection approaches in seven mental health RCTs and found both consistencies and inconsistencies in which variables were identified in each dataset across the different approaches. Some variables were identified consistently as important, some variables were identified consistently as unimportant, and other variables had mixed indications, depending on the variable selection method. We can have increased confidence in the importance of variables that are consistently identified as important, and similarly, that those variables rejected consistently should be considered unimportant. We also believe that we can use our understanding of the different methodologies to determine whether those variables that are identified in some approaches but not in others are inconsistently identified due to weak or noisy effects, and thus should be considered poor predictors, or whether this pattern can be attributed to shortcomings of specific approaches. For example, a variable might be selected by one approach that can flexibly model higher-order

interactions (Bleich et al. 2014) but excluded by a second that cannot (Austin & Tu 2004) if that variable's predictive relationship to outcome involves a three-way or nonlinear interaction.

Once the variables that have prognostic or prescriptive relationships to outcome have been identified, the model weights are specified. Model weights determine how much, and in what direction, each variable contributes to the prediction of outcome. Although the specifics of how a modeling approach characterizes these relationships can differ (e.g., parametric approaches, which might use linear regression, versus nonparametric machine-learning approaches, which might utilize tree-based modeling approaches), any of these approaches can generate predictions for new clients for each treatment condition for which the prediction is to be made. Both variable selection and weight setting should be performed using techniques that maximize the stability and generalizability of the model (Gillan & Whelan 2017).

EVALUATING TREATMENT RECOMMENDATION APPROACHES

As described in previous sections, once a model is built it can be used to generate predictions for each patient's outcomes. The utility of the model can then be evaluated on the basis of comparisons of the predictions with observed outcomes. This can be done either within the dataset that was used to generate the predictions or with a new sample of clients who receive (ideally, randomly) treatment A or treatment B. When the same dataset is used both to generate the model and to test its utility, special care must be taken to avoid a situation in which the model is fit specifically to the sample and is therefore unlikely to generalize in an independent application (Ioannidis 2005, Open Science Collaboration 2015).

To estimate the expected utility of the prediction-based recommendations without bias, data from the to-be-predicted patient cannot be included in the course of development of the algorithm (Hastie et al. 2009). This can be accomplished in model development with the use of bootstrapping or internal cross-validation methods. Ongoing efforts to refine feature selection and weight setting with cross-validation focus on ways of identifying robust feature sets and robust means of determining the weights that will be applied to those features. Well-constructed models are built with the aim of avoiding both underfitting and overfitting at both the feature-selection and weight-setting stages. The procedures for maximizing power (avoiding underfitting) and generalizability (by avoiding overfitting) are in continuous development.

Although there are many ways one could test a PAI prospectively, the most straightforward approach would be to randomize a new sample of clients to each treatment. A test of the utility of the model can then be derived from a comparison of the outcomes of those individuals who happen to be randomized to the intervention that was identified by the model as more likely to have a positive outcome, versus the outcomes of those who get randomized to their nonindicated intervention. In the context of equivalent average outcomes for the two treatments, if the average response of those who receive their indicated treatment is (statistically significantly) superior to the average response of those who receive their nonindicated treatment, this can be taken as evidence that the model has predictive power. Further examination of the size of this benefit in the context of other relevant factors (e.g., cost of administering the required assessments) would inform a judgment concerning the clinical utility of a model (Huang et al. 2012, 2015). Another approach to a prospective study would be to randomize participants to allocation-as-usual (AAU; for example, patient preference or clinical judgment) versus model-guided allocation. Although attractive for its comparison to a real-world treatment allocation strategy, this approach reduces the sample size available for comparison, as the only patients that can be used to compare the utility of the model are those for whom the AAU and model-based assignments disagree.

Careful consideration of the distinction between the different patient types reviewed earlier is important when evaluating treatment selection models. Indexes such as the PAI yield binary

recommendations (A versus B), but they also contain information about the strength of the recommendation. When used to inform treatment selection in the context of two treatments with equivalent average effects, many individuals can be expected to have PAIs close to zero, indicating that little to no difference in outcomes is predicted between the treatments. For these individuals, one implication is that either treatment could be recommended, as would be so for a type-3 pliant patient from **Figure 2**, who will respond to any treatment according to its strength. However, an individual with a PAI near zero might instead be a spontaneous remitter (type 1), an easy patient (type 2), a difficult patient (type 5), or an intractable patient (type 6). An examination of the within-treatment prognostic predictions will provide an indication of which profile best describes such an individual. Predictions of roughly equally poor outcomes in both treatments might indicate a challenging or intractable patient, whereas predictions of full symptom resolution in both treatments might indicate a spontaneous remitter, an easy patient, or a type-3 pliant patient. A patient with poor predicted outcomes in both treatments under consideration would tentatively be categorized as intractable (type 6), but it is possible that such a patient (type 5) might benefit from a treatment not included in the comparison, such as the combination of the two treatments studied. Identifying these individuals and recommending a stronger treatment could reduce the number of exposures to ineffective treatments.

A recommendation that treatment A is to be preferred over treatment B could arise from a PAI that is very large, in which case a clinician might strongly advise a client to pursue treatment A. However, if the predicted advantage is so small as to be clinically meaningless (e.g., a PAI close to zero), then the clinician would communicate this information to the client, and other factors would play a larger role in selecting treatment. Evidence for the importance of attending to recommendation strength can be found in the results of contexts in which greater expected benefit of treatment selection was observed for individuals with larger PAIs compared to those whose PAIs were smaller (Cohen et al. 2017, DeRubeis et al. 2014a, Huibers et al. 2015, Keefe et al. 2018).

DISCUSSION

Clinical practitioners and researchers have long sought knowledge about what works for whom. This knowledge matters. Many stakeholders would benefit from improvements in our ability to identify, for each individual, the intervention among those under consideration that is most likely to yield the best response. The implications for individuals are obvious. People suffering from depression want interventions that will work. Limiting the number of individuals exposed to ineffective first-line treatments and reducing the average time to recovery will not only reduce suffering from the symptoms of depression, but will increase economic productivity inasmuch as symptoms of depression interfere with a person's ability to perform work functions at a high level (Layard et al. 2007). Intelligent allocation of limited or costly resources has relevance for any class of treatment, including psychotherapy—the availability of which is often limited by the availability of trained clinicians—and pharmacotherapy, in which associated risks should be minimized.

Success in efforts to match individuals to treatments has been elusive. Historical attempts to use research findings to promote propitious matches of clients to treatments have relied on analyses that take into account a single feature of the client. Work with single features has been attractive in part due to its simplicity, and because of the ease with which a theory-based interpretation can be applied to the findings to support or understand the resulting recommendations. Unfortunately, the vast majority of this research on individual differences in treatment response (e.g., project MATCH; Allen et al. 1997) has failed to have a meaningful impact on client care (Simon & Perlis 2010).

Modern multivariable treatment selection approaches can overcome many of the shortcomings that have hindered progress and therefore hold great promise for the future of precision mental health. Part of this future will require a resolution of the tension between the statistical methodology of explanatory approaches that have dominated psychology and the predictive approaches the will power precision medicine going forward (Yarkoni & Westfall 2017).

Although it was not the focus of this review, we want to emphasize that we believe the treatment selection process should be an open and shared decision-making process between patients and clinicians. Treatment selection tools should be viewed as providing useful information that helps inform this collaborative decision-making process.

FUTURE DIRECTIONS

Research that informs treatment selection will continue to include analyses of data from RCTs, but it should and likely will also be conducted with large treatment databases (Kessler 2018), collected online or through electronic medical records (Perlis et al. 2012). The designs of RCTs will also be better tuned to the goals of precision mental health. Recent work has demonstrated the potential for dynamic assessment in precision mental health (Fernandez et al. 2017, Fisher & Boswell 2016). Modular psychotherapies that can be accessed online are fertile grounds for future efforts to personalize treatment for depression (Watkins et al. 2016). The pretreatment assessments that provide grist for treatment selection models will include biomarkers and other measures that promise to reveal prescriptive relationships, in addition to the self-report, demographic and clinical variables that have fueled most treatment selection findings reported to date. There are several ongoing studies, designed specifically to generate knowledge relevant to outcome prediction in depression treatment, that feature potential biomarkers, including information from neuroimaging and genetic testing (Brunoni et al. 2015, Lam et al. 2016, Williams et al. 2011). Two such trials are the Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care study (Trivedi et al. 2016), which focuses on two antidepressants (sertraline and bupropion) in the context of early-onset, recurrent MDD, and the recently completed Predictors of Remission in Depression to Individual and Combined Treatments study, which compared CBT to ADM in treatment-naive adults with moderate to severe MDD (Dunlop et al. 2012a). Lutz et al. (2017) recently initiated an RCT that tests personalized psychotherapy prediction and adaptation tools in a real-world clinic. The exploratory nature of many of the existing prediction models increases the importance of external validation and tests of generalizability. To realize the promise of precision mental health, existing models as well as those that are being developed will need to be validated prospectively against standard allocation schemes (Kingslake et al. 2017). Moreover, it will be important for all stakeholders, including providers and patients, to be involved in shaping the tools that will translate the findings into practice.

SUMMARY POINTS

1. Outcomes in depression could be improved by using patient characteristics that are associated with differential responses to treatment to help individuals select the right intervention.
2. Statistical models that predict treatment response can be constructed to generate individualized treatment recommendations.
3. Translating the growing interest in precision medicine into clinical support tools will require statistical approaches such as the Personalized Advantage Index and related efforts.

4. The implementation of precision medicine approaches has the potential to increase the efficiency of mental health systems, even without improving treatments or developing new ones, by optimizing the allocation of existing resources.

FUTURE ISSUES

1. Most existing prediction models have not been validated on independent samples.
2. Clinical trials comparing treatment selection to alternative methods of allocation (e.g., clinical judgment or patient preference) should be undertaken to evaluate the clinical value of implementing actuarial treatment recommendations.
3. Novel statistical approaches (e.g., machine learning) have the potential to generate more powerful prediction models and improved treatment recommendations.
4. As the cost and feasibility of collecting biological predictors from sources such as neuroimaging and genetic assays diminish, the ability of these variables to improve treatment selection models should be explored.
5. Most efforts to model the differential prediction of outcome have relied on data from clinical trials with relatively small samples, whereas future research should also address the potential of large electronic databases from health care systems to inform precision mental health.
6. System-level factors, as well as the needs and preferences of clinicians and patients, will need to be incorporated into the design of treatment decision tools to ensure that they are acceptable, ethical, and easy to implement.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors acknowledge with gratitude the extensive and challenging feedback on previous drafts of this article by Sona Dimidjian, Yoni Ashar, Rob Saunders, Paul Andrews, and Steve Hollon. We thank Thomas Kim for his help preparing the manuscript, including his assistance with the literature reviews that informed **Supplemental Table 1**. We are grateful to the MQ Foundation, whose support has made this work possible. Finally, we acknowledge the extensive network of colleagues and collaborators who have contributed to this emerging area. To witness the eagerness of members of this network to work together and to share ideas, expertise and data, with the common goal of improving outcomes in mental health, has been most gratifying.

LITERATURE CITED

Allen J, Mattson M, Miller W, Tonigan J, Connors G, et al. 1997. Matching alcoholism treatments to client heterogeneity. *J. Stud. Alcohol* 58:7–29

American Psychiatric Association. 2010. *Practice Guideline for the Treatment of Patients with Major Depressive Disorder*. Arlington, VA: Am. Psych. Assoc. Pub. 3rd ed. **https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf**

Amsterdam JD, Lorenzo-Luaces L, DeRubeis RJ. 2016. Step-wise loss of antidepressant effectiveness with repeated antidepressant trials in bipolar II depression. *Bipolar Disord.* 18:563–70

Amsterdam JD, Shults J. 2009. Does tachyphylaxis occur after repeated antidepressant exposure in patients with Bipolar II major depressive episode? *J. Affect. Disord.* 115:234–40

Amsterdam JD, Williams D, Michelson D, Adler LA, Dunner DL, et al. 2009. Tachyphylaxis after repeated antidepressant drug exposure in patients with recurrent major depressive disorder. *Neuropsychobiology* 59:227–33

Ashar YK, Chang LJ, Wager TD. 2017. Brain mechanisms of the placebo effect: an affective appraisal account. *Annu. Rev. Clin. Psychol.* 13:73–98

Austin PC, Tu JV. 2004. Bootstrap methods for developing predictive models. *Am. Stat.* 58:131–37

Barber JP, Muenz LR. 1996. The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: empirical findings from the Treatment for Depression Collaborative Research Program. *J. Consult. Clin. Psychol.* 64:951–58

Barbui C, Cipriani A, Patel V, Ayuso-Mateos JL, van Ommeren M. 2011. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. *Br. J. Psychiatry* 198:11–16

Baron RM, Kenny DA. 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51:1173–82

Beck AT, Rush AJ, Shaw BF, Emery G. 1979. *Cognitive Therapy of Depression*. New York: Guilford Press

Beutler LE, Engle D, Mohr D, Daldrup RJ, Bergan J, et al. 1991. Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *J. Consult. Clin. Psychol.* 59:333–40

Bleich J, Kapelner A, George EI, Jensen ST. 2014. Variable selection for BART: an application to gene regulation. *Ann. Appl. Stat.* 8:1750–81

Bossuyt PM, Parvin T. 2015. Evaluating biomarkers for guiding treatment decisions. *EJIFCC* 26:63–70

Brunoni AR, Sampaio-Junior B, Moffa AH, Borrione L, Nogueira BS, et al. 2015. The Escitalopram versus Electric Current Therapy for Treating Depression Clinical Study (ELECT-TDCS): rationale and study design of a non-inferiority, triple-arm, placebo-controlled clinical trial. *São Paulo Med. J.* 133:252–63

Bursac Z, Gauss CH, Williams DK, Hosmer DW. 2008. Purposeful selection of variables in logistic regression. *Source Code Biol. Med.* 3:17

Byar DP. 1985. Assessing apparent treatment—covariate interactions in randomized clinical trials. *Stat. Med.* 4:255–63

Byar DP, Corle DK. 1977. Selecting optimal treatment in clinical trials using covariate information. *J. Chronic Diseas.* 30:445–59

Byrne SE, Rothschild AJ. 1998. Loss of antidepressant efficacy during maintenance therapy: possible mechanisms and treatments. *J. Clin. Psychiatry* 59:279–88

Chakraborty B, Moodie E. 2013. *Statistical Methods for Dynamic Treatment Regimes*. New York: Springer

Chambless DL, Hollon SD. 1998. Defining empirically supported therapies. *J. Consult. Clin. Psychol.* 66:7–18

Cheavens JS, Strunk DR, Lazarus SA, Goldstein LA. 2012. The compensation and capitalization models: a test of two approaches to individualizing the treatment of depression. *Behav. Res. Ther.* 50:699–706

Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G. 2017. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA Psychiatry* 74:370–78

Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, et al. 2016. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 3:243–50

Cloitre M, Petkova E, Su Z, Weiss B. 2016. Patient characteristics as a moderator of post-traumatic stress disorder treatment outcome: combining symptom burden and strengths. *BJPsych Open* 2:101–06

Cohen Z, Kim T, Van R, Dekker J, Driessen E. 2017. Individual treatment recommendations of cognitive-behavioral or psychodynamic therapy for mild to moderate adult depression: improving the Personalized Advantage Index approach. **https://osf.io/6qxve/**

Craske MG, Meuret AE, Ritz T, Treanor M, Dour HJ. 2016. Treatment for anhedonia: a neuroscience driven approach. *Depress. Anxiety* 33:927–38

An early example of a prescriptive multivariable treatment selection model in medicine.

Cronbach LJ. 1957. The two disciplines of scientific psychology. *Am. Psychol.* 12:671–84
Cuijpers P, Huibers MJ, Furukawa TA. 2017. The need for research on treatments of chronic depression. *JAMA Psychiatry* 74:242–43
d'Agostino RB. 1998. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat. Med.* 17:2265–81
Dawes RM. 1979. The robust beauty of improper linear models in decision making. *Am. Psychol.* 34:571–82
Dawes RM. 2005. The ethical implications of Paul Meehl's work on comparing clinical versus actuarial prediction methods. *J. Clin. Psychol.* 61:1245–55
Dawes RM, Faust D, Meehl PE. 1989. Clinical versus actuarial judgment. *Science* 243:1668–74
Delgadillo J, Huey D, Bennett H, McMillan D. 2017. Case complexity as a guide for psychological treatment selection. *J. Consult. Clin. Psychol.* 85:835–53
Delgadillo J, Moreea O, Lutz W. 2016. Different people respond differently to therapy: a demonstration using patient profiling and risk stratification. *Behav. Res. Ther.* 79:15–22
DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. 2014a. The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLOS ONE* 9:e83875
DeRubeis RJ, Gelfand LA, German RE, Fournier JC, Forand NR. 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.* 24:419–28
DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, et al. 2005. Cognitive therapy versus medications in the treatment of moderate to severe depression. *Arch. Gen. Psychiatry* 62:409–16
Doove LL, Dusseldorp E, Van Deun K, Van Mechelen I. 2014. A comparison of five recursive partitioning methods to find person subgroups involved in meaningful treatment-subgroup interactions. *Adv. Data Anal. Classification* 8:403–25
Driessen E, Cuijpers P, Hollon SD, Dekker JJ. 2010. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *J. Consult. Clin. Psychol.* 78:668–80
Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, et al. 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* 23:28–38
Dunlop BW, Binder EB, Cubells JF, Goodman MM, Kelley ME, et al. 2012a. Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. *Trials* 13:106
Dunlop BW, Kelley ME, Aponte-Rivera V, Mletzko-Crowe T, Kinkead B, et al. 2017. Effects of patient preferences on outcomes in the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) Study. *Am. J. Psychiatry* 174:546–56
Dunlop BW, Kelley ME, Mletzko TC, Velasquez CM, Craighead WE, Mayberg HS. 2012b. Depression beliefs, treatment preference, and outcomes in a randomized trial for major depressive disorder. *J. Psychiatr. Res.* 46:375–81
Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, et al. 1989. National Institute of Mental Health treatment of depression collaborative research program: general effectiveness of treatments. *Arch. Gen. Psychiatry* 46:971–82
Fernandez KC, Fisher AJ, Chi C. 2017. Development and initial implementation of the Dynamic Assessment Treatment Algorithm (DATA). *PLOS ONE* 12:e0178806
Fineberg NA, Brown A, Reghunandanan S, Pampaloni I. 2012. Evidence-based pharmacotherapy of obsessive-compulsive disorder. *Int. J. Neuropsychopharmacol.* 15:1173–91
Fisher AJ, Boswell JF. 2016. Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment* 23:496–506
Forand NR, Huibers MJ, DeRubeis RJ. 2017. Prognosis moderates the engagement–outcome relationship in unguided cCBT for depression: a proof of concept for the prognosis moderation hypothesis. *J. Consult. Clin. Psychol.* 85:471–83
Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, et al. 2010. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 303:47–53

Early article reviewing the superiority of and resistance to statistical decision making.

The first published presentation of the Personalized Advantage Index (PAI) approach to treatment selection.

A simulation study demonstrating the importance of patient types in mental health treatment research.

Fournier JC, DeRubeis RJ, Shelton RC, Gallop R, Amsterdam JD, Hollon SD. 2008. Antidepressant medications v. cognitive therapy in people with depression with or without personality disorder. *Br. J. Psychiatry* 192:124–29

Fournier JC, DeRubeis RJ, Shelton RC, Hollon SD, Amsterdam JD, Gallop R. 2009. Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *J. Consult. Clin. Psychol.* 77:775–85

Gabrieli JD, Ghosh SS, Whitfield-Gabrieli S. 2015. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85:11–26

Gail M, Simon R. 1985. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* 41:361–72

Garge NR, Bobashev G, Eggleston B. 2013. Random forest methodology for model-based recursive partitioning: the mobForest package for R. *BMC Bioinformat.* 14:125

Gillan CM, Daw ND. 2016. Taking psychiatry research online. *Neuron* 91:19–23

Gillan CM, Whelan R. 2017. What big data can do for treatment in psychiatry. *Curr. Opin. Behav. Sci.* 18:34–42

Gordon E, Rush AJ, Palmer DM, Braund TA, Rekshan W. 2015. Toward an online cognitive and emotional battery to predict treatment remission in depression. *Neuropsychiatr. Dis. Treat.* 11:517–31

Green KC, Armstrong JS. 2015. Simple versus complex forecasting: the evidence. *J. Bus. Res.* 68:1678–85

Grove WM, Meehl PE. 1996. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy. *Psychol. Public Policy Law* 2:293–323

Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. 2000. Clinical versus mechanical prediction: a meta-analysis. *Psychol. Assess.* 12:19–30

Gunn J, Wachtler C, Fletcher S, Davidson S, Mihalopoulos C, et al. 2017. Target-D: a stratified individually randomized controlled trial of the diamond clinical prediction tool to triage and target treatment for depressive symptoms in general practice: study protocol for a randomized controlled trial. *Trials* 18:342

Gunter L, Zhu J, Murphy S. 2011a. Variable selection for qualitative interactions. *Stat. Methodol.* 8:42–55

Gunter L, Zhu J, Murphy S. 2011b. Variable selection for qualitative interactions in personalized medicine while controlling the family-wise error rate. *J. Biopharm. Stat.* 21:1063–78

Hamburg MA, Collins FS. 2010. The path to personalized medicine. *N. Engl. J. Med.* 2010:301–4

Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer. 2nd ed.

Hingorani AD, van der Windt DA, Riley RD, Abrams K, Moons KG, et al. 2013. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 346:e5793

Hollon SD, Areán PA, Craske MG, Crawford KA, Kivlahan DR, et al. 2014. Development of clinical practice guidelines. *Annu. Rev. Clin. Psychol.* 10:213–41

Hollon SD, Thase ME, Markowitz JC. 2002. Treatment and prevention of depression. *Psychol. Sci. Public Interest* 3:39–77

Holmes EA, Craske MG, Graybiel AM. 2014. A call for mental-health science. *Nature* 511:287–89

Howland RH. 2014. Pharmacogenetic testing in psychiatry: not (quite) ready for primetime. *J. Psychosoc. Nurs. Ment. Health Serv.* 52:13–16

Huang Y, Gilbert PB, Janes H. 2012. Assessing treatment-selection markers using a potential outcomes framework. *Biometrics* 68:687–96

Huang Y, Laber EB, Janes H. 2015. Characterizing expected benefits of biomarkers in treatment selection. *Biostatistics* 16:383–99

Huibers MJ, Cohen ZD, Lemmens LH, Arntz A, Peeters FP, et al. 2015. Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the Personalized Advantage Index Approach. *PLOS ONE* 10:e0140771

Hunter AM, Cook IA, Greenwald S, Tran ML, Miyamoto KN, Leuchter AF. 2011. The Antidepressant Treatment Response (ATR) index and treatment outcomes in a placebo-controlled trial of fluoxetine. *J. Clin. Neurophysiol.* 28:478–82

Imai K, Ratkovic M. 2013. Estimating treatment effect heterogeneity in randomized program evaluation. *Ann. Appl. Stat.* 7:443–70

Review of the potential for prediction, big data, and machine learning to advance psychiatry.

Meta-analysis finding that actuarial/mechanical/statistical prediction is consistently superior to clinical prediction.

Iniesta R, Malki K, Maier W, Rietschel M, Mors O, et al. 2016a. Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *J. Psychiatr. Res.* 78:94–102

Iniesta R, Stahl D, McGuffin P. 2016b. Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol. Med.* 46:2455–65

Ioannidis JP. 2005. Why most published research findings are false. *PLOS Med.* 2:e124

James G, Witten D, Hastie T, Tibshirani R. 2013. *An Introduction to Statistical Learning*. New York: Springer

Jamshidian M, Jalal S. 2010. Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika* 75:649–74

Janes H, Pepe MS, Bossuyt PM, Barlow WE. 2011. Measuring the performance of markers for guiding treatment decisions. *Ann. Intern. Med.* 154:253–59

Jollans L, Whelan R. 2016. The clinical added value of imaging: a perspective from outcome prediction. *Biol. Psychiatry: Cogn. Neurosci. Neuroimaging* 1:423–32

Kapelner A, Bleich J. 2016. bartMachine: A powerful tool for machine learning. *J. Stat. Softw.* 70:1–40

Katsnelson A. 2013. Momentum grows to make "personalized" medicine more "precise." *Nat. Med.* 19:249

Keefe JR, Wiltsey-Stirman S, Cohen ZD, DeRubeis RJ, Smith BN, Resick P. 2018. In rape-trauma PTSD, patient characteristics indicate which trauma-focused treatment they are most likely to complete. *Depression Anxiety.* In press

Kessler RC. 2018. The potential of predictive analytics to provide clinical decision support in depression treatment planning. *Curr. Opin. Psychiatry* 31:32–39

Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, et al. 2016. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol. Psychiatry* 21:1366–71

Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, et al. 2017. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiol. Psychiatr. Sci.* 26:22–36

Khan A, Leventhal RM, Khan SR, Brown WA. 2002. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J. Clin. Psychopharmacol.* 22:40–45

King M, Walker C, Levy G, Bottomley C, Royston P, Weich S. 2008. Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study. *Arch. Gen. Psychiatry* 65:1368–76

Kingslake J, Dias R, Dawson GR, Simon J, Goodwin GM, et al. 2017. The effects of using the PReDicT Test to guide the antidepressant treatment of depressed patients: study protocol for a randomised controlled trial. *Trials* 18:558

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. 2008. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLOS Med.* 5:e45

Klerman GL, Weissman MM. 1994. *Interpersonal Psychotherapy of Depression: A Brief, Focused, Specific Strategy*. Lanham, MD: Jason Aronson, Inc.

Kocsis JH, Leon AC, Markowitz JC, Manber R, Arnow B, et al. 2009. Patient preference as a moderator of outcome for chronic forms of major depressive disorder treated with nefazodone, cognitive behavioral analysis system of psychotherapy, or their combination. *J. Clin. Psychiatry* 70:354–61

Kraemer HC. 2013. Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. *Stat. Med.* 32:1964–73

Kraemer HC, Blasey CM. 2004. Centring in regression analyses: a strategy to prevent errors in statistical inference. *Int. J. Methods Psychiatr. Res.* 13:141–51

Excellent resource for those interested in building statistical prediction models.

Kuhn M, Johnson K. 2013. *Applied Predictive Modeling*. New York: Springer

Lam RW, Milev R, Rotzinger S, Andreazza AC, Blier P, et al. 2016. Discovering biomarkers for antidepressant response: protocol from the Canadian Biomarker Integration Network in Depression (CAN-BIND) and clinical characteristics of the first patient cohort. *BMC Psychiatry* 16:105

Layard R, Clark D, Knapp M, Mayraz G. 2007. Cost-benefit analysis of psychological therapy. *Natl. Inst. Econ. Rev.* 202:90–98

Leuchter AF, Cook IA, Gilmer WS, Marangell LB, Burgoyne KS, et al. 2009. Effectiveness of a quantitative electroencephalographic biomarker for predicting differential response or remission with escitalopram and bupropion in major depressive disorder. *Psychiatry Res.* 169:132–38

Leykin Y, Amsterdam JD, DeRubeis RJ, Gallop R, Shelton RC, Hollon SD. 2007a. Progressive resistance to a selective serotonin reuptake inhibitor but not to cognitive therapy in the treatment of major depression. *J. Consult. Clin. Psychol.* 75:267

Leykin Y, DeRubeis RJ, Gallop R, Amsterdam JD, Shelton RC, Hollon SD. 2007b. The relation of patients' treatment preferences to outcome in a randomized clinical trial. *Behavior. Ther.* 38:209–17

Lo A, Chernoff H, Zheng T, Lo S-H. 2015. Why significant variables aren't automatically good predictors. *Proc. Natl. Acad. Sci.* 112:13892–97

Lorenzo-Luaces L, DeRubeis RJ, Bennett IM. 2015. Primary care physicians' selection of low-intensity treatments for patients with depression. *Fam. Med.* 47:511–16

Lorenzo-Luaces L, DeRubeis RJ, 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

Luedtke AR, van der Laan MJ. 2016. Super-learning of an optimal dynamic treatment rule. *Int. J. Biostat.* 12:305–32

Lutz W, Hofmann SG, Rubel J, Boswell JF, Shear MK, et al. 2014. Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *J. Consult. Clin. Psychol.* 82:287

Lutz W, Saunders SM, Leon SC, Martinovich Z, Kosfelder J, et al. 2006. Empirically and clinically useful decision making in psychotherapy: differential predictions with treatment response models. *Psychol. Assess.* 18:133–41

Lutz W, Zimmermann D, Müller VN, Deisenhofer A-K, Rubel JA. 2017. Randomized controlled trial to evaluate the effects of personalized prediction and adaptation tools on treatment outcome in outpatient psychotherapy: study protocol. *BMC Psychiatry* 17:306

Ma J, Stingo FC, Hobbs BP. 2016. Bayesian predictive modeling for genomic based personalized treatment selection. *Biometrics* 72:575–83

MacKinnon DP, Fairchild AJ, Fritz MS. 2007. Mediation analysis. *Annu. Rev. Psychol.* 58:593–614

Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, et al. 2005. Deep brain stimulation for treatment-resistant depression. *Neuron* 45:651–60

McCullough JP Jr. 2003. Treatment for chronic depression: cognitive behavioral analysis system of psychotherapy (CBASP). *J. Pschother. Integr.* 13:241–63

McGirr A, Berlim M, Bond D, Fleck M, Yatham L, Lam R. 2015. A systematic review and meta-analysis of randomized, double-blind, placebo-controlled trials of ketamine in the rapid treatment of major depressive episodes. *Psychol. Med.* 45:693–704

McGrath CL, Kelley ME, Holtzheimer PE, Dunlop BW, Craighead WE, et al. 2013. Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry* 70:821–29

McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. 2013. Patient preference for psychological versus pharmacologic treatment of psychiatric disorders: a meta-analytic review. *J. Clin. Psychiatry* 74:595–602

Meehl PE. 1954. *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Washington, DC: Am. Psychol. Assoc. http://psycnet.apa.org/record/2006-21565-000

Seminal monograph of Paul Meehl's original lectures in which he championed statistical decision making.

Meehl PE. 1978. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *J. Consult. Clin. Psychol.* 46:806–34

Mergl R, Henkel V, Allgaier AK, Kramer D, Hautzinger M, et al. 2011. Are treatment preferences relevant in response to serotonergic antidepressants and cognitive-behavioral therapy in depressed primary care patients? Results from a randomized controlled trial including a patients' choice arm. *Psychother. Psychosom.* 80:39–47

Mickey RM, Greenland S. 1989. The impact of confounder selection criteria on effect estimation. *Am. J. Epidemiol.* 129:125–37

National Health Service. 2016. *Psychological Therapies, Annual Report on the Use of IAPT Services: England 2015–16*. London, UK: Health Soc. Care Inf. Cent.

National Institute for Health and Clinical Excellence. 2009. *Depression: Treatment and Management of Depression in Adults*. London: Br. Psychol. Soc.

National Research Council. 2011. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC: Natl. Acad. Press

Report summarizing recent progress in precision medicine, including compelling examples of treatment selection in cancer.

Nemeroff CB, Heim CM, Thase ME, Klein DN, Rush AJ, et al. 2003. Differential responses to psychotherapy versus pharmacotherapy in patients with chronic forms of major depression and childhood trauma. *Proc. Natl. Acad. Sci.* 100:14293–96

Nigatu YT, Liu Y, Wang J. 2016. External validation of the international risk prediction algorithm for major depressive episode in the US general population: the PredictD-US study. *BMC Psychiatry* 16:256

Niles AN, Loerinc AG, Krull JL, Roy-Byrne P, Sullivan G, et al. 2017a. Advancing personalized medicine: application of a novel statistical method to identify treatment moderators in the Coordinated Anxiety Learning and Management Study. *Behav. Ther.* 48:490–500

Niles AN, Wolitzky-Taylor KB, Arch JJ, Craske MG. 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

Nuzzo R. 2014. Scientific method: statistical errors. *Nature* 506:150–52

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349:aac4716

Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, et al. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304:1497–500

Pao W, Miller VA. 2005. Epidermal growth factor receptor mutations, small-molecule kinase inhibitors, and non-small-cell lung cancer: current knowledge and future directions. *J. Clin. Oncol.* 23:2556–68

Papakostas GI, Fava M. 2010. *Pharmacotherapy for Depression and Treatment-Resistant Depression*. Singapore: World Scientific

Parmigiani G. 2002. *Modeling in Medical Decision Making: A Bayesian Approach*. Chichester, UK: Wiley

Passos IC, Mwangi B, Kapczinski F. 2016. Big data analytics and machine learning: 2015 and beyond. *Lancet Psychiatry* 3:13–15

Pauker SG, Kassirer JP. 1980. The threshold approach to clinical decision making. *N. Engl. J. Med.* 302:1109–17

Paul GL. 1967. Strategy of outcome research in psychotherapy. *J. Consult. Psychol.* 31:109–18

Perlis R, Iosifescu D, Castro V, Murphy S, Gainer V, et al. 2012. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychol. Med.* 42:41–50

Perlis RH. 2013. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol. Psychiatry* 74:7–14

Perlis RH. 2016. Abandoning personalization to get to precision in the pharmacotherapy of depression. *World Psychiatry* 15:228–35

Perlis RH, Fijal B, Dharia S, Heinloth AN, Houston JP. 2010. Failure to replicate genetic associations with antidepressant treatment response in duloxetine-treated patients. *Biol. Psychiatry* 67:1110–13

Perlis RH, Patrick A, Smoller JW, Wang PS. 2009. When is pharmacogenetic testing for antidepressant response ready for the clinic? A cost-effectiveness analysis based on data from the STAR*D study. *Neuropsychopharmacology* 34:2227–36

Petkova E, Ogden RT, Tarpey T, Ciarleglio A, Jiang B, et al. 2017. Statistical analysis plan for stage 1 EMBARC (Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care) study. *Contemporary Clin. Trials Commun.* 6:22–30

Pizzagalli DA. 2011. Frontocingulate dysfunction in depression: toward biomarkers of treatment response. *Neuropsychopharmacology* 36:183–206

Preference Collaborative Review Group. 2008. Patients' preferences within randomised trials: systematic review and patient level meta-analysis. *BMJ* 337:a1864

Raza GT, Holohan DR. 2015. Clinical treatment selection for posttraumatic stress disorder: suggestions for researchers and clinical trainers. *Psychol. Trauma* 7:547–54

Renjilian DA, Perri MG, Nezu AM, McKelvey WF, Shermer RL, Anton SD. 2001. Individual versus group therapy for obesity: effects of matching participants to their treatment preferences. *J. Consult. Clin. Psychol.* 69:717–21

Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, et al. 2012. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* 13:239–46

Rubenstein L, Rayburn N, Keeler E, Ford D, Rost K, Sherbourne C. 2007. Predicting outcomes of primary care patients with major depression: development of a depression prognosis index. *Psychiatr. Serv.* 58:1049–56

Rush AJ, Trivedi MH, Stewart JW, Nierenberg AA, Fava M, et al. 2011. Combining Medications to Enhance Depression Outcomes (CO-MED): acute and long-term outcomes of a single-blind randomized study. *Am. J. Psychiatry* 168:689–701

Saunders R, Cape J, Fearon P, Pilling S. 2016. Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. *J. Affect. Disord.* 197:107–15

Schleidgen S, Klingler C, Bertram T, Rogowski WH, Marckmann G. 2013. What is personalized medicine: sharpening a vague term based on a systematic literature review. *BMC Med. Ethics* 14:55

Schneider RL, Arch JJ, Wolitzky-Taylor KB. 2015. The state of personalized treatment for anxiety disorders: a systematic review of treatment moderators. *Clin. Psychol. Rev.* 38:39–54

Schwaederle M, Zhao M, Lee JJ, Eggermont AM, Schilsky RL, et al. 2015. Impact of precision medicine in diverse cancers: a meta-analysis of phase II clinical trials. *J. Clin. Oncol.* 33:3817–25

Simon GE, Perlis RH. 2010. Personalized medicine for depression: can we match patients with treatments? *Am. J. Psychiatry* 167:1445–55

Smagula SF, Wallace ML, Anderson SJ, Karp JF, Lenze EJ, et al. 2016. Combining moderators to identify clinical profiles of patients who will, and will not, benefit from aripiprazole augmentation for treatment resistant late-life major depressive disorder. *J. Psychiatr. Res.* 81:112–18

Stephan KE, Schlagenhauf F, Huys QJ, Raman S, Aponte EA, et al. 2017. Computational neuroimaging strategies for single patient predictions. *Neuroimage* 145:180–99

Steyerberg E. 2008. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer

Comprehensive resource for building and evaluating prediction models for clinical applications.

Swift JK, Callahan JL. 2009. The impact of client treatment preferences on outcome: a meta-analysis. *J. Clin. Psychol.* 65:368–81

Swift JK, Callahan JL, Vollmer BM. 2011. Preferences. *J. Clin. Psychol.* 67(2):155–65

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58:267–88

Tiemens B, Bocker K, Kloos M. 2016. Prediction of treatment outcome in daily generalized mental health care practice: first steps towards personalized treatment by clinical decision support. *Eur. J. Pers. Cent. Healthcare* 4:24–32

Trivedi MH, McGrath PJ, Fava M, Parsey RV, Kurian BT, et al. 2016. Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. *J. Psychiatr. Res.* 78:11–23

Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, et al. 2006. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am. J. Psychiatry* 163:28–40

Tversky A, Kahneman D. 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90:293–315

Uher R, Huezo-Diaz P, Perroud N, Smith R, Rietschel M, et al. 2009. Genetic predictors of response to antidepressants in the GENDEP project. *Pharmacogenom. J.* 9:225–33

Uher R, Perlis R, Henigsberg N, Zobel A, Rietschel M, et al. 2012. Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychol. Med.* 42:967–80

Uher R, Tansey KE, Dew T, Maier W, Mors O, et al. 2014. An inflammatory biomarker as a differential predictor of outcome of depression treatment with escitalopram and nortriptyline. *Am. J. Psychiatry* 171:1278–86

van Straten A, Tiemens B, Hakkaart L, Nolen W, Donker M. 2006. Stepped care versus matched care for mood and anxiety disorders: a randomized trial in routine practice. *Acta Psychiatr. Scand.* 113:468–76

Vittengl JR, Clark LA, Thase ME, Jarrett RB. 2017. Initial steps to inform selection of continuation cognitive therapy or fluoxetine for higher risk responders to cognitive therapy for recurrent major depressive disorder. *Psychiatry Res.* 253:174–81

Vittengl JR, Jarrett RB, Weitz E, Hollon SD, Twisk J, et al. 2016. Divergent outcomes in cognitive-behavioral therapy and pharmacotherapy for adult depression. *Am. J. Psychiatry* 173:481–90

Wallace ML, Frank E, Kraemer HC. 2013. A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *JAMA Psychiatry* 70:1241–47

Wang R, Ware JH. 2013. Detecting moderator effects using subgroup analyses. *Prev. Sci.* 14:111–20

Wasserstein RL, Lazar NA. 2016. The ASA's statement on p-values: context, process, and purpose. *Am. Stat.* 70:129–33

Watkins E, Newbold A, Tester-Jones M, Javaid M, Cadman J, et al. 2016. Implementing multifactorial psychotherapy research in online virtual environments (IMPROVE-2): study protocol for a phase III trial of the MOST randomized component selection method for internet cognitive-behavioural therapy for depression. *BMC Psychiatry* 16:345

Weisz JR, Krumholz LS, Santucci L, Thomassin K, Ng MY. 2015. Shrinking the gap between research and practice: tailoring and testing youth psychotherapies in clinical care contexts. *Annu. Rev. Clin. Psychol.* 11:139–63

Weitz ES, Hollon SD, Twisk J, van Straten A, Huibers MJ, et al. 2015. Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy versus pharmacotherapy: an individual patient data meta-analysis. *JAMA Psychiatry* 72:1102–9

Wellek S. 1997. Testing for absence of qualitative interactions between risk factors and treatment effects. *Biometrical J.* 39:809–21

Westover AN, Kashner TM, Winhusen TM, Golden RM, Nakonezny PA, et al. 2015. A systematic approach to subgroup analyses in a smoking cessation trial. *Am. J. Drug Alcohol Abuse* 41:498–507

Widaman KF, Helm JL, Castro-Schilo L, Pluess M, Stallings MC, Belsky J. 2012. Distinguishing ordinal and disordinal interactions. *Psychol. Methods* 17:615–22

Williams LM, Rush AJ, Koslow SH, Wisniewski SR, Cooper NJ, et al. 2011. International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: rationale and protocol. *Trials* 12:4

Winter SE, Barber JP. 2013. Should treatment for depression be based more on patient preference? *Patient Pref. Adher.* 7:1047–57

World Health Organization. 2017. *Depression and other common mental disorders: global health estimates*. Rep. CC BY-NC-SA 3.0 IGO, World Health Org., Geneva

Yakovlev AY, Goot RE, Osipova TT. 1994. The choice of cancer treatment based on covariate information. *Stat. Med.* 13:1575–81

Yarkoni T, Westfall J. 2017. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect. Psychol. Sci.* 12:1100–22

Zilcha-Mano S, Keefe JR, Chui H, Rubin A, Barrett MS, Barber JP. 2016. Reducing dropout in treatment for depression: translating dropout predictors into individualized treatment recommendations. *J. Clin. Psychiatry* 77:e1584–e90

Zimmerman M, Clark HL, Multach MD, Walsh E, Rosenstein LK, Gazarian D. 2015. Have treatment studies of depression become even less generalizable? A review of the inclusion and exclusion criteria used in placebo-controlled antidepressant efficacy trials published during the past 20 years. *Mayo Clin. Proc.* 90:1180–86

Zimmerman M, Clark HL, Multach MD, Walsh E, Rosenstein LK, Gazarian D. 2016. Symptom severity and the generalizability of antidepressant efficacy trials: changes during the past 20 years. *J. Clin. Psychopharmacol.* 36:153–56

Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc.: Ser. B* 67:301–20

Annual Review of
Clinical Psychology

Volume 14, 2018

Contents

Errata

An online log of corrections to *Annual Review of Clinical Psychology* articles may be found at http://www.annualreviews.org/errata/clinpsy