CHAPTER 7

Moving Beyond "One Size Fits All"

ZACHARY D. COHEN
YONI K. ASHAR
ROBERT J. DERUBEIS

Half a century ago, Gordon Paul (1967) reviewed psychotherapy outcome research and concluded that "in 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?" (p. 111, original emphasis). Paul's article has been extensively cited, and the previous passage has often been invoked or abbreviated to "What works for whom?" The idea is a good one, recognizing that no single approach is likely to be best for every person who presents with a mental health problem.

Improving outcomes across mental health treatments is imperative (Holmes et al., 2018). The predominant approach has been to develop new treatments, such as novel neurological (e.g., deep brain stimulation: Mayberg et al., 2005), pharmacological (e.g., ketamine: McGirr et al., 2015), and psychological treatments (e.g., positive affect treatment: Craske, Meuret, Ritz, Treanor, & Dour, 2016). However, this approach on its own is likely to be insufficient. Depression, which is the world's leading cause of disability (World Health Organization, 2017), provides an example. Numerous evidence-based interventions for major depressive disorder (MDD) have been developed, but the average treatment response rate remains at about 50% (Luty et al., 2007; National Health Service, 2016; Papakostas & Fava, 2010). In this chapter we describe an alternative approach—*treatment selection,* which aims to identify for each person the treatment that is best for him or her.

Treatment selection focuses on the role of "individual differences"—the idea that individuals respond differently to treatment, and that these differences can be studied and characterized. It responds directly to the principle that therapists "attend to the individual person to make the complex choices necessary to conceptualize, prioritize, and treat multiple symptoms," as described by the American Psychological Association (2006, p. 279) and as highlighted by other evidence-based practice models (see Spring, Hoffman, & Steglitz, Chapter 1, this volume).

The treatment selection approach has been applied profitably in medicine (Ashley, 2016; National Research Council, 2011), where it has been termed *precision* medicine (Hamburg & Collins, 2010) or *personalized* medicine (Katsnelson, 2013; Schleidgen, Klinger, Bertram, Rogowski, & Marckmann, 2013). For example, in oncology, the choice of anticancer drug can be tailored to match genetic mutations detected in a particular patient's tumor, leading to improved outcomes (Paez et al., 2004; Rosell et al., 2012). Could similar approaches help improve outcomes in mental health? This is the question that motivates this chapter.

Traditionally, results from clinical trials have been analyzed and reported in terms of *average treatment response across all individuals*. In a typical study, researchers compare responses to treatments *A, B,* and *C,* and declare the treatment with the greatest average response the "winner." This approach has been described as a "horse race" approach to psychotherapy research. And, in lumping patients together, it obscures the very information that could best guide personalized clinical decision making.

In reality, each treatment produces a wide distribution of treatment responses. In treatment selection, it is the distribution of response scores that is of interest rather than the average treatment response. To analyze and report treatment effects at the level of individual patients, rather than treatment-average effects, new methods are needed. The third section of this chapter provides an accessible introduction to these emerging methods.

In the next section, we describe treatment selection as practiced by mental health clinicians today. In doing so, we review the empirical support for the kinds of treatment selection decisions clinicians make every day.

Treatment Selection as Practiced by Clinicians Today

Most clinicians seek to move beyond the "one size fits all" approach and to prioritize the *person* sitting across the room. They realize the importance of recognizing differences in diagnostic presentations and the specific problems with which clients struggle, as well as the importance of the client's personality and other sociodemographic and cultural considerations. For example, a client with social anxiety may be struggling primarily with low self-esteem, poor social skills, a history of scarring social interactions, or a combination of these factors. A client also might present with a range of comorbid diagnoses, such as substance abuse problems or eating disorders, all of which might influence clinical decision making. Clinicians also frequently prioritize the problems the client most wants to address. For example, a depressed woman might want to focus on interpersonal issues or, alternatively, on how to make progress toward employment goals. Most clinicians also attend to features of the client's environment relevant to treatment. A young man who is homeless and estranged from his parents might require a different approach than would a father with three children who is stressed at work and in his marriage.

A clinician who attends to information about a specific client's presentation will generate hypotheses about the client's expected response to a given treatment (Lorenzo-Luaces, De Rubeis, & Bennett, 2015; Raza & Holohan, 2015). Here is an example from one of our intake reports in our training clinic (modified to protect the client's anonymity), which exemplifies such a line of reasoning:

> Given the client's difficulty with emotion tolerance and impulsivity, and in light of her history of self-injury and dissociation, the imaginal exposure interventions in prolonged exposure for posttraumatic stress disorder (PTSD) might prove especially difficult at this time. To minimize the risk of self-harm, it will be important to provide interventions that target emotion regulation and coping skills to help her engage productively, with a subsequent focus on imaginal exposure.

Such a conclusion may draw on a variety of sources, including a clinician's history with clients with similar features, his or her experiences in training and supervision, reasoning based on theory, and the empirical literature on treatment response (Cook, Dinnen, Simiola, Thompson, & Schnurr, 2014; Raza & Holohan, 2015). In fact, gathering and integrating such relevant information to inform recommendations about treatment are key tasks that clinicians do every day.

Unfortunately, the gap that has existed between clinical practice and clinical research has left clinicians with limited empirical guidance for this process. As a result, most clinicians (ourselves included) have tended to use an approach consistent with what Perlis (2016) has dubbed *artisanal medicine*.

Artisanal and Actuarial Medicine

Artisanal medicine refers to 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 the empirical evidence. Given the historical paucity of statistical approaches that could inform clinical decision making, artisanal approaches are typically the only option available for tailoring treatment to the individual.

Artisanal approaches are hobbled by several limitations that limit the validity and utility of

such approaches for decision making (Dawes, 1979, 2005; Dawes, Faust, & Meehl, 1989; Perlis, 2016; Tversky & Kahneman, 1983). For example, when a clinician notes improvements in a patient's life, it is tempting to conclude that the improvement resulted from treatment, when it in fact could have been driven by factors in the patient's life unrelated to treatment. Hannan and colleagues (2005) examined clinicians' ability to predict their own clients' response to treatment and found that clinicians only predicted deterioration in .01% of their clients, in contrast to the 7.3% who actually deteriorated. Clinicians are also unreliable when assessing their own skills and outcomes: For example, when a large sample of mental health professionals were asked to compare their own clinical skills and performance to those of their peers, 25% indicated their skill was at the 90th percentile or higher, and none viewed themselves as being below average (Walfish, McAlister, O'Donnell, & Lambert, 2012). Clinicians also overestimated their clients' rates of improvement and underestimated their rates of deterioration (Walfish et al., 2012). A meta-analysis of 75 studies on clinician judgment accuracy revealed that clinicians with more experience or education were only modestly more accurate in their predictions compared to less experienced clinicians (Spengler et al., 2009). The factors, findings, and examples we described earlier do not mean that clinicians are especially bad at those kinds of judgments. It simply means they are human. The limitations of human judgment, which influence all of us, have been well described by Tversky and Kahneman (1974) and by Lilienfeld and colleagues (Lilienfeld, Ammirati, & David, 2012; Lilienfeld, Ammirati, & Landfield, 2009).

Actuarial decision making—defined as making predictions in a statistical, algorithmic, and reproducible way (Grove, Zald, Lebow, Snitz, & Nelson, 2000)—can overcome some of the difficulties inherent in human judgment (e.g., Dawes et al., 1989; Pauker & Kassirer, 1980). Research on this approach suggests that decision making via actuarial processes is more efficient than clinical judgment (Grove & Meehl, 1996). In practice, clinicians rarely are able to observe the counterfactuals that would be needed to assess the validity of their decisions. Consider the following example: If a clinician believes that a given client would be better off in psychodynamic therapy than in behavioral activation therapy, the clinician will likely attempt to ensure that the patient receives psychodynamic therapy. If the clinician is successful, then the only outcome that can be observed for that client is how he or she actually fares in psychodynamic therapy. The clinician does not have the opportunity to learn how the client *would have done* had he or she received behavioral activation; thus, the clinician is missing crucial information to evaluate the validity of his or her judgment. Given the complexity of each patient, it is unlikely that any one clinician has the opportunity to treat and observe the outcomes of enough similar patients to construct a valid decision rule. Of course, it is possible that some clinicians possess implicit models to guide treatment selection that produce better predictions than others; however, there have been few (if any) studies identifying and characterizing these individuals or their implicit models. Given the lack of data regarding clinicians who are "good" at treatment allocation, new clinicians cannot be trained in this important ability.

The case for using actuarial methods to personalize treatment in mental health was made forcefully over 60 years ago by Paul Meehl (1954). The field of mental health treatment has only just begun to apply Meehl's line of thinking. Some treatment selection decisions in clinics today do use features of actuarial decision making, in that they use measurable variables that have been the focus of empirical research. Clinicians who value evidence-based practice (EBP) can make treatment selection decisions guided by data regarding (1) diagnosis, (2) the client's treatment preferences, (3) the client's self-reported response to previous treatments, and (4) symptom severity. How good is the evidence base that guides such decisions? We answer this question in the following sections.

Using Diagnosis to Guide Treatment Selection

For much of the 20th century, scientific efforts focused on characterizing the core pathologies of the DSM-defined diagnostic categories and developing related treatments. This was the most relevant research available for a clinician who sought empirical guidance in tailoring treatment to the client sitting in front of him or her. Successes in these efforts included evidence for specific treatments for specific disorders (e.g., cognitive-behavioral therapy [CBT] for MDD (Beck, Rush, Shaw, & Emery, 1979). These interventions have been evaluated in randomized

clinical trials (RCTs; see Kraemer & Periyakoil, Chapter 4, this volume), in which active treatments have been compared to control conditions or to other active treatments. Based on findings from such studies, these treatments have been considered to be "empirically supported" for clients with the associated diagnosis (Chambless & Hollon, 1998). Similarly, specific classes of psychiatric drugs have been investigated under the assumption that they are best suited for specific disorders (Fineberg, Brown, Reghunandanan, & Pampaloni, 2012). Thus, clinicians were provided a resource for tailoring treatments by attending to the diagnostic status of the client.

It makes sense that treatment recommendations would follow from accurate diagnosis. For example, antibiotic treatment is appropriate for bacterial but not viral infection. The same is true in some mental health contexts. For example, CBT for bulimia nervosa has been shown to be superior to other forms of psychotherapy (Linardon, Wade, de la Piedad Garcia, & Brennan, 2017). Thus, a clinician working with a patient whose primary problem is bulimia nervosa has a clear first-line treatment recommendation. Similarly, when treating a patient with the diagnosis of bipolar disorder, clinical practice guidelines, based on substantial evidence, recommend mood stabilizers (e.g., lithium) or second-generation antipsychotics (Connolly & Thase, 2011). However, for many mental health diagnoses, especially depression, there exist an abundance of empirically supported treatments (ESTs) that have roughly similar efficacy. Thus, a depression diagnosis is not particularly informative for treatment selection. Moreover, many clients present with more than one disorder (Hirschfeld, 2001; Kessler, Chiu, Demler, & Walters, 2005; Kircanski, LeMoult, Ordaz, & Gotlib, 2017), requiring clinicians to use the artisanal approach to sequence and combine different ESTs for clients. In summary, for some conditions (e.g., bulimia nervosa, mania), a clear diagnosis guides a clear treatment recommendation. But for many other conditions and comorbidities, diagnosis often cannot point us toward a preferred treatment.

Using Client Preference to Guide Treatment Selection

EBP guidelines also specify the importance of attending to clients' preferences. This recommendation is based on not only respect for the client's autonomy and dignity but also an assumption that a treatment preferred by a client will outperform a nonpreferred treatment. Surprisingly, studies examining the relationship between client preference and treatment outcomes include findings that are positive (Kocsis et al., 2009; Mergl et al., 2011; Swift & Callahan, 2009; Swift, Callahan, & Vollmer, 2011), mixed (Dunlop et al., 2017; McHugh, Whitton, Peckham, Welge, & Otto, 2013; Preference Collaborative Review Group, 2008), and even negative (Dunlop et al., 2012, Leykin, DeRubeis, et al., 2007; Renjilian et al., 2001; Winter & Barber, 2013). One reason for this is that a patient in an RCT, by virtue of agreeing to randomization, is indicating that he or she does not have a *strong* preference for one treatment or another. Another reason is that a patient may not have a good sense of what to expect in a particular treatment, resulting in a relatively uninformed preference. More research is needed on patient preferences in typical clinical (nonrandomized) treatment contexts and on the value of providing patients with more information about treatment options prior to decision making. In summary, although respecting a patient's preference is an integral part of ethical clinical practice, its utility for guiding what will work best for an individual patient is unclear.

Using Previous Treatment Experience to Guide Treatment Selection

Numerous outcome studies have found that treatment history is associated with future response for pharmacological treatments. For example, prior exposure to and history of nonresponse to antidepressant medications (ADMs) have each been found consistently to predict poor outcome to future courses of antidepressants (Amsterdam, Lorenzo-Luaces, & DeRubeis, 2016; Amsterdam & Shults, 2009; Amsterdam et al., 2009; Byrne & Rothschild, 1998). Moreover, there is evidence that the number of prior ADM exposures predicts response differentially across ADM and cognitive therapy (CT): Leykin, Amsterdam, and colleagues (2007) found that multiple previous ADM-exposures predicted a poorer response to ADM, but not to CT, such that patients with two or more prior exposures to ADMs were more likely to benefit from CT than from ADM. Clearly, assessing pharmacological treatment history is important and could be used to inform treatment selection.

Very little research exists on the relationship between prior psychotherapy and future response to treatment (Boswell, McAleavey,

Castonguay, Hayes, & Locke, 2012). In addition to the difficulty of accurately assessing prior psychotherapy, this research is complicated by the correlation between factors that are independently associated with reduced likelihood of response to treatment (e.g., recurrent, chronic, and treatment-resistant forms of depression) and treatment history. Grenyer, Deane, and Lewis (2008) found no relationship between prior psychotherapy and response to supportive–expressive dynamic psychotherapy for depression. However, Boswell and colleagues (2012) found that prior psychotherapy (as well as prior psychotropic medication) were associated with decreased response to counseling. These findings stand in contrast to recent work by Blau and DiMino (2018), who found that college students with prior counseling experience had more favorable outcomes relative to never-counseled students. Additional research is needed in this area, with increased focus on assessing the type and dosage of prior psychotherapy, and on the distinction between prior positive response and prior exposure. At this point, the utility of this information for guiding what will work best for an individual patient is unclear.

Using Symptom Severity to Guide Treatment Selection

Many clinicians assign patients with less severe symptoms to less intensive treatments, and patients with more severe symptoms to more intensive treatments (Lorenzo-Luaces et al., 2015). There is evidence to support this practice. Active treatments (e.g., ADMs, CBT) have greater efficacy relative to control treatments (e.g., pill-placebo, psychological control treatment) at higher levels of pretreatment depression severity, indicating that more severely depressed patients need more intensive treatment (Barbui, Cipriani, Patel, Ayuso-Mateos, & van Ommeren, 2011; Driessen, Cuijpers, Hollon, & Dekker, 2010; Fournier et al., 2010; Khan, Leventhal, Khan, & Brown, 2002; Kirsch et al., 2008). Similarly, most practice guidelines use symptom severity as an indicator that stronger treatments or combination treatments (e.g., ADMs and psychotherapy) are preferred over lower-intensity interventions (American Psychiatric Association, 2010; National Institute for Health and Clinical Excellence, 2009). However, in comparisons of two active treatments of similar intensity levels for major depression, baseline severity does not moderate treatment outcomes (Vittengl et al., 2016; Weitz et al., 2015).

In contrast, a common view among clinicians who work with patients diagnosed with PTSD is that stronger treatments (e.g., trauma-focused CBTs such as prolonged exposure [PE] or cognitive processing therapy [CPT]) are contraindicated for more patients with more severe symptoms or more complex presentations. Consistent with this view, there is some evidence that the superiority of stronger over weaker treatments in PTSD is greater among patients whose presentations are less severe (Wiltsey Stirman et al., 2019). Many clinicians see complex patients as being "unready" to engage with more intensive interventions (Cook et al., 2014; Rosen et al., 2016). Empirical support for this belief, which is reinforced by practice guidelines, has been mixed (Cook, Simiola, Hamblen, Bernardy, & Schnurr, 2017; Osei-Bonsu et al., 2017). Thus, at the present moment, the literature does not provide a clear treatment recommendation for patients with PTSD with higher levels of symptom severity.

Relatedly, many clinicians assume that patients with higher symptom severity need medication and not psychotherapy (*www.webmd.com/depression/guide/understanding-depression-treatment#3*), a belief that is reinforced by some practice guidelines (*http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf*). Studies, reviews, and meta-analyses have not supported such assumptions belief: Psychotherapy and ADMs are equally effective across the range of baseline severity (DeRubeis et al., 2005; Furukawa et al., 2017; Simon & Perlis, 2010; Weitz et al., 2015), at least in outpatient contexts.

Summary

The existing evidence does not provide a clear guide for clinicians who value EBP. Data are available to support the use of diagnosis, client preference, treatment history, and symptom severity for some clients but not for others. For this reason, the field has moved to focus more heavily on multivariable approaches to treatment selection.

Interpreting Results of Treatment Selection Research

In this section, we provide an overview of treatment selection methodology, with an eye toward helping the reader interpret findings reported in

the treatment selection literature. We first describe the difference between prescriptive and prognostic variables, then discuss the interpretation of treatment selection research results.

Prescriptive and Prognostic Variables

Treatment selection approaches largely[1] rely upon variables that can be measured prior to treatment and that predict treatment outcomes reliably. Such variables can be considered as either prescriptive or prognostic. *Prescriptive variables* indicate which treatment is better for a patient, among several treatment options. Prescriptive variables have often been referred to as *moderators* in the research literature. They affect the direction or strength of the differences in outcome between two or more treatments (Baron & Kenny, 1986). For example, Fournier and colleagues (2008) found that the presence of a comorbid personality disorder among depressed patients predicted better response to ADM relative to CT, while the absence of a personality disorder predicted a better response to CT than to ADM.

A variable is *prognostic* if it predicts response regardless of which treatment is delivered. For example, higher baseline depression severity is a prognostic variable: It is associated with worse outcomes, for both medications and CT (Weitz et al., 2015). Prognostic information can be used to provide realistic expectations to the treating clinician, as well as the client and family, regardless of the treatment chosen. Additionally, prognostic variables can identify patients who are unlikely to respond, and such patients can be monitored more closely (Lutz et al., 2014). Prognostic variables do not provide the kind of information that is needed to optimize the choice between two or more treatment options, as they indicate the responsiveness of the patient to treatment in general.

Whether a variable is prognostic or prescriptive can depend on the context of treatment options being considered. For example, baseline depression severity predicts outcomes similarly for CT and medication treatments (Weitz et al., 2015), making it *prognostic* in this context. However, higher baseline severity predicts a larger advantage of medication over placebo and of psychotherapy over nondirective supportive counseling, making baseline severity *prescriptive* in this context (Ashar, Chang, & Wager, 2017; Driessen et al., 2010; Fournier et al., 2010).

Interpreting Prognostic Variables

A common misinterpretation of a prognostic finding is to infer that clients found to have a poor prognosis in a given treatment will fare better with a different treatment (Simon & Perlis, 2010). For example, consider the finding that in CT, patients with chronic depression have lower recovery rates than those with nonchronic depression (Fournier et al., 2009). One might be tempted to conclude that other interventions such as ADM treatment or psychological treatments specifically targeting chronic depression (e.g., CBASP; McCullough, 2003) should be preferred to CT for individuals with chronic depression. Alternatively, it could be that CT is as effective as other available treatments for chronic depression (Cuijpers, Huibers, & Furukawa, 2017). This alternative hypothesis is supported by an RCT that compared CT to ADM and found that chronicity was *prognostic*: It was associated with similarly lower response rates in both treatments (Fournier et al., 2009). Likewise, an RCT comparing CBASP to ADM in individuals with chronic depressions found no difference in response rates (Nemeroff et al., 2003). Thus, prognostic findings do not necessarily provide good information on which to base a preference between two or more evidence-based treatments.[2]

Interpreting Prescriptive Variables

Prescriptive variables can be difficult to interpret correctly. In this section, we focus on single prescriptive variables, often called *moderators,* which have been the main focus of research to date. Moderator research has been attractive in part because of the relative simplicity of research designs and statistical analyses when

[1] Some proposed treatment selection approaches rely on "early response" indicators. These approaches cannot answer the question about which treatment to recommend at the initiation of treatment and thus are beyond the focus of this chapter.

[2] Two prognostic models validated within the same sample could be used together to make treatment recommendations for an individual. For example, one model would predict response to treatment *A,* the other model would predict response to treatment *B,* and the two predictions would be compared to select a treatment. This approach was proposed by Kessler and colleagues (2017) and adapted by Deisenhofer and colleagues (2018) for the purpose of guiding treatment decisions in PTSD.

analyzing a single variable. However, making sense of findings from such studies and applying results to clinical practice is often not straightforward.

Consider the following scenario. A clinician is presented with a depressed patient and must choose between CBT or ADM treatment. The clinician finds a paper that concludes that "clients with a greater number of prior ADM exposures fare worse in ADM treatment." Should the clinician use this finding to select a treatment for the patient?

In fact, this simple description of the research finding is insufficient to inform a decision, since it is consistent with several different patterns of relationship between prior ADM exposures and treatment outcomes. We depict six hypothetical relationships in Figure 7.1. Critically, without knowing which of these six is the true relationship, a treatment recommendation cannot be made.

If the data are as depicted in Figure 7.1a, the clinician should prescribe ADMs only if the client has had 0 or 1 prior ADM exposures. If data are as depicted in Figure 7.1b, the clinician should opt for CT for all patients except those with no prior ADM exposures, for whom no difference is predicted between CT and ADM. If the data are as depicted in Figure 7.1f, either CT or ADM is indicated because clients with prior ADM exposures fare worse in both treatments. In fact, the findings on which the statement is based is depicted in Figure 7.1e (Leykin, Amsterdam, et al., 2007).

This example illustrates one of the many ways in which the same statement—"clients with a greater number of prior ADM exposures fared worse in ADM treatment"—can refer to importantly different patterns, which is why such statements by themselves are insufficiently detailed to inform treatment selection decisions.

In empirical reports, the distinctions are rarely made between the different types of prescriptive relationships depicted in Figure 7.1. When the details of these relations are only implied rather than precisely stated, they lead to inconsistent, misleading, or simply incorrect interpretations. We offer this as a warning to readers wading into the treatment moderator literature.

The Source of the Data

In addition to paying careful attention to the specific results, it is critical to consider the source of the data. To date, most prediction have utilized data from randomized trials. This has inherent limitations when attempting to generalize to the typical (nonrandomized) treatment context. Future efforts, exemplified by the ongoing work of Gillan and Daw (2016) to collect mental health treatment outcome data online also incorporate naturalistic (nonrandomized) data (Kessler, 2018), such as large treatment databases that can be generated from electronic medical records (EMR; Perlis et al., 2012). However, the potential influence of unknown confounds (i.e., third variable problems) is a limitation of treatment selection efforts outside the context of RCT data. The bias in predictions in such studies can derive from the "selection effects" that result when clients with a given feature (e.g., history of nonresponse to ADMs) are preferentially provided a given treatment (e.g., CT). A full discussion of the potential risks of using nonrandomized data is beyond the scope of this review, but a recent analysis by Agniel, Kohane, and Weber (2018) highlights the complexity of such efforts. The authors reviewed hospital EMR data from more than 600,000 patients on 272 common laboratory tests and found that for 68% of the tests, the *timing* of when the labs were ordered were more predictive of patient survival than the results of the tests. Despite these obstacles, the promise of EMR or electronic health record (EHR) data is compelling; Simon and colleagues (2018) used demographic and clinical EHR data from 3 million patients seen in primary care and mental health specialty clinics to develop a 90-day suicide risk index. Their model relied on information about prior suicide attempts, mental health and substance use diagnoses, responses to the suicide question from a commonly used depression questionnaire, and prior inpatient or emergency mental health treatment. Individuals in the top 5% of their risk index accounted for almost half of all subsequent suicide attempts and deaths by suicide. Predictive tools of this kind have great promise in helping to inform and improve clinical decision making.

Another key issue in interpreting findings rests with characteristics of the research sample and research treatment. It is risky to generalize research findings to a population outside the one from which the research sample was drawn. For example, findings regarding one treatment may not hold for a treatment believed to be similar, as evidenced by many reports that sets of variables and models found to predict treatment response to one ADM have failed to generalize to a different ADM (Chekroud et al., 2016; Ini-

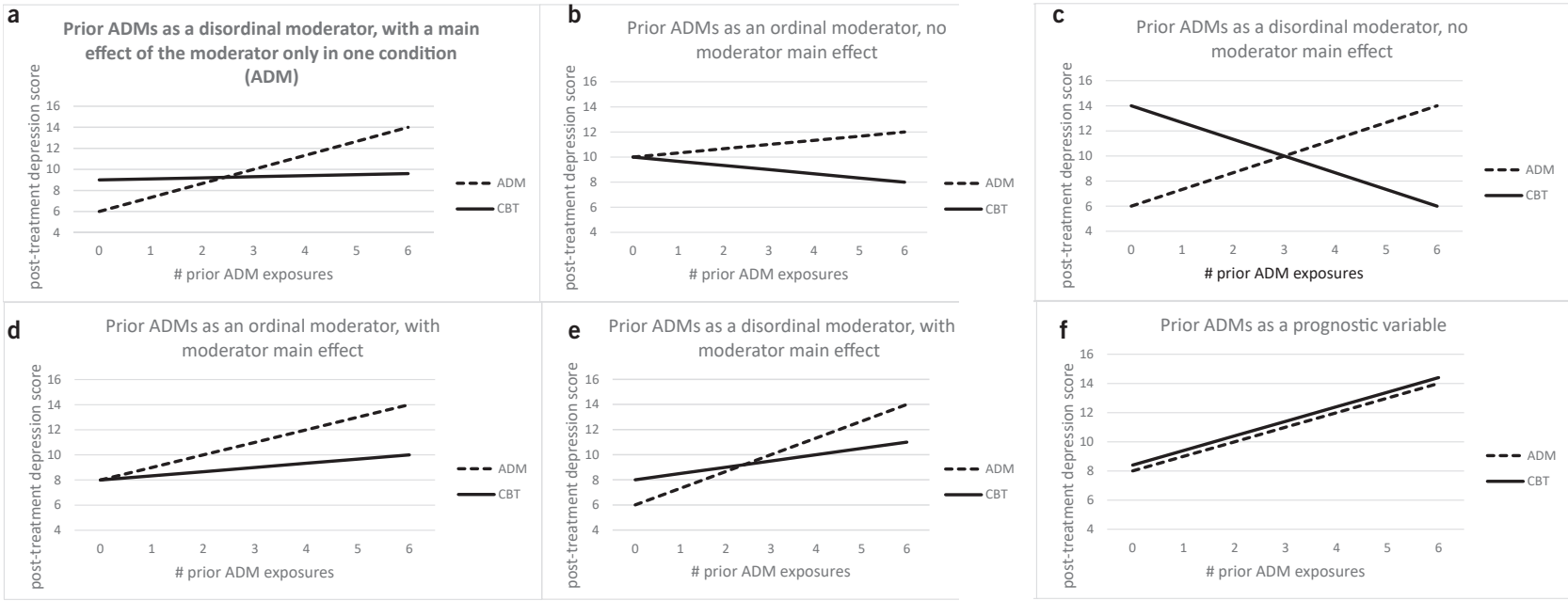

FIGURE 7.1. Six hypothetical relationships that could support the finding "clients with a greater number of prior ADM (antidepressant) exposures fared worse in ADM treatment." The dashed line represents expected posttreatment depression score in ADM, and the solid line represents expected outcome in cognitive-behavioral therapy (CBT).

esta et al., 2018; Iniesta, Malki, et al., 2016; Perlis, Fijal, Dharia, Heinloth, & Houston, 2010). Clinicians seeking to apply research findings must consider to what extent they can have confidence that (1) the client in front of them comes from the same population as that in which the research was conducted and (2) the treatment being considered is similar to the treatment delivered in the study.

Finally, a finding is only a finding until it is replicated. Statistical relationships reported from a lone research investigation should be used with great caution. This is especially true of findings in the prediction literature (relative to, for example, findings from experimental manipulations), as these often arise from exploratory analyses that are at higher risk for false positives.

Statistical and Clinical Significance

Reliance on tests of significance can result in misleading impressions about the importance of predictive variables (Nuzzo, 2014; Wasserstein & Lazar, 2016). This can happen in at least two ways. First, statistical significance testing typically applies arbitrary thresholds (e.g., $p < .05$) to determine which variables matter. However, the difference in the predictive utility of an excluded variable that "just missed" the threshold (e.g., $p = .06$) and one that is "barely" significant (e.g., $p = .04$) is trivial (Mickey & Greenland, 1989). As Rosnow and Rosenthal (1989, p. 1277) noted: "Dichotomous significance testing has no ontological basis. . . . Surely, God loves the .06 nearly as much as the .05." Further complicating the matter is that most clinical trials do not have sufficient statistical power to detect moderators at conventional statistical thresholds (i.e., $p < .05$), such that moderators will not be detected unless they are unusually strong.

Second, statistically significant results should be interpreted in context of their clinical significance. In large sample sizes, statistically significant results associated with small effect sizes can be of negligible clinical significance (Lo, Chernoff, Zheng, & Lo, 2015; Meehl, 1978). If a variable predicts a difference of 0.2 points on an anxiety scale with treatment *A* instead of *B,* does this really matter?[3] Janes, Pepe, Bossuyt, and Barlow (2011) proposed a series of questions, none of which invoke statistical significance, to use when evaluating treatment selection markers:

1. Does the marker help patients choose amongst treatment options?
2. If the marker is measured as a continuous variable, how should information from it be used to inform treatment decisions?
3. What is the expected impact on the population of using the marker to select treatment?
4. For what proportion of patients is a change in the treatment recommendation likely if the marker is measured and patients' values on it are used in decision making?

Moving beyond statistics, consideration of factors such as cost, feasibility, and client burden should be weighed against the additive predictive power provided by variables that must be collected specially for treatment selection (Perlis, Patrick, Smoller, & Wang, 2009).

Furthermore, outcomes for evaluating the benefits of data-informed treatment selection should include more than reductions in the scores obtained on disorder-specific symptom questionnaires, although the vast majority of treatment selection models have been constructed using only such data (cf. Wallace, Frank, & Kraemer, 2013). Outcome indices that combine symptom change with measures of side effects, social and occupational functioning, and quality of life will support a more informative, holistic approach to treatment selection (Kraemer & Frank, 2010).

Multivariate Models for Treatment Selection

Development of the Personalized Advantage Index

The vast majority of research on the prediction of treatment response in mental health has focused on the role of one predictor, considered in isolation. This is understandable: If a single predictive variable associated with clinically meaningful differences can be identified in a treatment context, application to practice is likely to be straightforward (despite all the caveats and cautions we have described). Moreover, if a single variable can account for how well one treatment will perform relative to another for any given individual, it is likely that the variable reflects an important mechanism of one or both treatments. However, as we discuss

[3] Although small effects may not be clinically significant on the individual level, they can often have meaningful impacts when considered at the population level.

in more detail later in the chapter, this "univariate" approach has largely failed to generate powerful predictions and has therefore had little impact on mental health practice (Simon & Perlis, 2010). This has led researchers to pivot to multivariate predictive approaches in recent years (Cohen & DeRubeis, 2018). Multivariate models integrate information from *multiple* predictors jointly (e.g., age, severity, employment status, social support) to generate a treatment recommendation. Multivariable models are likely to yield more powerful predictions (Chekroud et al., 2016; Delgadillo, Huey, Bennett, & McMillan, 2017; Iniesta, Malki, et al., 2016; Koutsouleris et al., 2016; Kraemer, 2013; Perlis, 2013), and they comport with our understanding of psychopathology and treatment response as complex, multiply determined phenomena (Drysdale et al., 2017).

We began to explore the possibility that multivariable linear modeling or machine learning[4] (Iniesta, Stahl, & McGuffin, 2016; Passos, Mwangi, & Kapczinski, 2016) approaches could be brought to bear on precision medicine problems in mental health in 2011. This quest was initiated with a specific goal in mind: to find or develop an approach that could identify clients with MDD for whom ADMs are likely to be more beneficial than CT, and vice versa. Two of our own findings prompted this 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, five variables (marital status, employment status, personality disorder (PD) comorbidity, antidepressant treatment history, and the number of recent stressful life events) had been identified that independently served as moderators of symptom change in this sample (Fournier et al., 2009). However, none of these five variables were powerful enough to separate those for whom one of the treatments was likely to produce greater symptom change relative to the other (Fournier et al., 2009).

Moreover, the five variables were relatively uncorrelated with each other, suggesting that they reflected different dimensions of treatment response. The variables acted like vectors, such that a patient's value on one variable could point to a slight advantage for ADM, while his or her value on another variable might point in the direction of CT. For example, as noted earlier, clients who were unemployed improved more in CT than in ADM (Fournier et al., 2009). It was also the case that clients with comorbid PD improved more with ADM than they did in CT, whereas clients without comorbid PD improved more in CT than in ADM (Fournier et al., 2008). But what recommendation should be made for a client with comorbid PD (indicating ADM) who was unemployed (indicating CT)? How does the clinician integrate this conflicting information along with the other three variables when forming a treatment recommendation? Multivariable models are needed to recommend a treatment in such situations.

Another important concept was that effective guidance for clinicians and clients would be unlikely to be best described in binary terms. Instead, for some clients, the difference in outcome between treatments might be quite substantial, while for others it would be negligible, and for still others in between. To address these challenges, DeRubeis, Cohen, and colleagues (2014) developed the personalized advantage index (PAI) approach, which has been the foundation for recent efforts by several different research teams (Cohen, Kim, Van, Dekker, & Driessen, 2019; Deisenhofer et al., 2018; Huibers et al., 2015; Keefe et al., 2018; Vittengl, Clark, Thase, & Jarrett, 2017; Webb et al., 2018; Zilcha-Mano et al., 2016).

The PAI approach to treatment selection involves the generation of a prediction of the expected differential benefit (in graded terms) of one intervention over one or more alternative treatment options. This begins with the identification of pretreatment variables in a dataset that predict differential response to two or more treatments. Once these predictors (moderators) are identified, a multivariable statistical model is constructed, comprising main effects of predictors as well as interaction terms representing the prescriptive variables' interaction with treatment.[5] An individual's PAI derives from the difference between his or her predicted out-

[4] "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" (Gillan & Whelan, 2017, p. 35).

[5] 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. For examples of this approach, see Deisenhofer and colleagues (2018) or Schweizer and colleagues (2019).

comes in two treatments (T × A and T × B). The sign of the difference indicates which of the two treatments is expected to be preferred for that patient. The magnitude of the difference reflects the magnitude of the predicted advantage of the indicated treatment over the nonindicated treatment. If a PAI is large, one might strongly advise a client to pursue a specific treatment. However, if the predicted advantage is small (e.g., a PAI close to zero), then one's recommendation might be more tempered (Cohen & DeRubeis, 2018).

We also have extended our use of the PAI to characterize and utilize patient subtypes in treatment selection decision processes. Some patients do equally well or equally poorly in all treatments, while for others, the particular treatment matters. The prototype clients (adapted from Cohen & DeRubeis, 2018; DeRubeis, Gelfand, German, Fournier, & Forand, 2014) described in Figure 7.2 aim to help in better understanding the types of clients for whom treatment selection might be relevant. For easy patients, any level of active treatment (from the highest to lowest strength) would result in high levels of improvement. For the challenging patients at the other end of the spectrum, little to no improvement would be expected at any level of therapy strength. *Pliant patients* are defined as those whose improvement would vary as a function of therapy quality, such that with very poor quality therapy or no therapy, little to no improvement would be expected, and with the highest quality therapy possible, significant improvement would result. The pliant patient category may be broken down further into two subgroups: individuals who would improve if they received quality treatment of any type versus other individuals who would improve only if "matched" to the specific treatment they receive (see Figure 7.2, types 2 and 2a/2b). These latter individuals (type 2a/2b), who respond

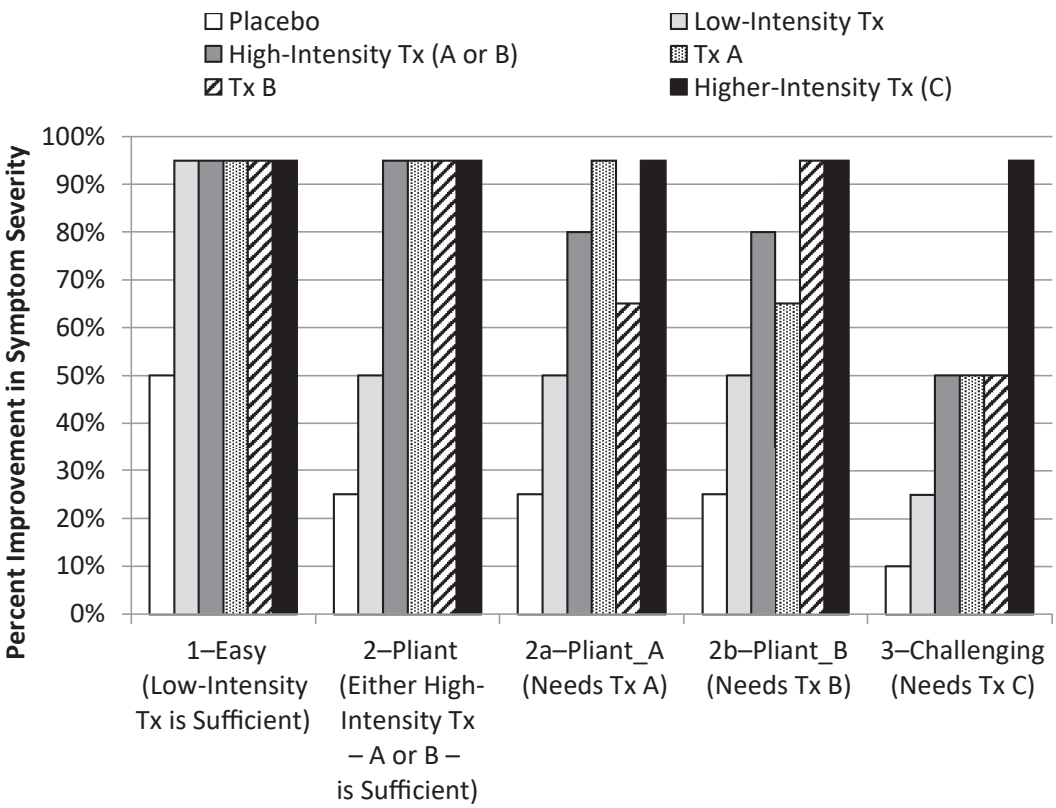

FIGURE 7.2. Depiction of expected improvement for five patient prototypes in different treatment contexts. *Note.* Tx, treatment.

well but only to a *specific* treatment, are the individuals for whom treatment selection will be most important.

This insight can be leveraged in a stratified care context. Patients who respond equally well or equally poorly to all treatments could be given the cheapest/safest treatment. However, ethical considerations arise in prescribing a weak treatment to a severely ill patient, simply because the patient is statistically unlikely to respond to a stronger treatment. On the other hand, "pliant" patients could be matched to the treatment of appropriate strength/cost/risk.

Lorenzo-Luaces, DeRubeis, van Straten, and Tiemens (2017) implemented such an approach as a proof of concept, with data from a randomized comparison of a high-intensity treatment (CT) with low-intensity treatment (brief therapy [BT]) and treatment as usual (TAU). On average, the differences between the high-intensity treatment and each of the two comparison conditions were small (van Straten, Tiemens, Hakkaart, Nolen, & Donker, 2006). The reasonable interpretation given by the authors was that high-intensity treatment was no more effective in this population than low-intensity treatment (Van Straten et al., 2006). However, when Lorenzo-Luaces and colleagues fit a multivariable model that arrayed patients from those with the poorest to the best prognoses, those with poor prognoses evidenced significantly higher rates of recovery in CBT (60%) than in BT (44%) or TAU (39%). In contrast, patients with model-predicted better prognoses evidenced nearly identical recovery rates across the three treatments. These findings represent an example of how treatment selection principles may be applied in stratified medicine in a way that might increase substantially the efficiency of mental health treatment systems.

The approach used to construct PAI models described earlier can be adapted to inform stratified medicine decisions for stepped care, where the choice is often between a high- versus low-intensity treatment. When the high-intensity treatment is more effective, on average, the goal is usually to differentiate between individuals who are likely to benefit much more from the high-intensity treatment than the low-intensity treatment and individuals for whom the expected differential benefit is small. In this case, patients can be arrayed along a continuum, at one end of which are patients who would be expected to evidence a poor response to a weak treatment, but who might benefit substantially from a strong treatment, especially if it is one that is suited to them. At the other end of the continuum are patients who are predicted to experience a positive response irrespective of the quality of the treatment. These are patients for whom a minimal treatment is expected to produce as much or nearly as much benefit as a strong, intensive treatment.

Other Multivariable Prescriptive Approaches

Other approaches to multivariate modeling also have been proposed in the mental health context. Barber and Muenz's (1996) reanalysis of data from the Treatment of Depression Collaborative Research Program (TDCRP; Elkin et al., 1989) provided one of the earliest examples of multivariable prediction of treatment response in mental health. Using data from the comparison of CT to interpersonal psychotherapy (IPT) for MDD, the authors built a "matching factor" that combined the prescriptive value of several moderators (marital status, avoidance, obsessiveness, and baseline severity) in a linear regression model predicting symptom change. The authors also explored the prescriptive value of two PD diagnoses, avoidant PD and obsessive–compulsive PD, and proposed that models including these factors could be used to match patients to CT or IPT.

Since this early work, many different analytic approaches have been taken to develop multivariate models. Drawing on a novel statistical technique, Lutz and colleagues (2006) used the "nearest neighbor" modeling approach to predict differential outcomes between two variations of CBT. This approach was adapted from methods used to predict avalanches in the Swiss Alps. Each client's outcome in each treatment is predicted from the outcomes of a group of clients who are most similar to the index client, and who have undergone either of the treatments. Here, similarity is defined by calculating the Euclidean distance between two clients' values on a vector of factors (i.e., the square root of the sum of the squared differences in values for each of the standardized variables). Within this group of "neighbors," the average outcome in each of the two interventions is calculated. These averages are considered the best predictions of outcome for the index client in each of the two treatments.

Kraemer (2013) proposed a statistical approach to treatment selection that involves the

creation of a single variable (termed *M**, and most often referred to as *M Star*) that represents a weighted combination of multiple moderators. Using data from a randomized comparison of IPT versus the antidepressant escitalopram, Wallace and colleagues (2013) demonstrated the approach, creating a combined moderator that comprised eight predictors: baseline depression severity, psychomotor activation, medical reassurance, number of depressive episodes, age, gender, anxiety, employment status. Recently, other groups have used the *M** approach to analyze studies of treatment-resistant late-life depression (Smagula et al., 2016) and anxiety disorders (Niles, Loerinc, et al., 2017; Niles, Wolitzky-Taylor, Arch, & Craske, 2017).

The evolution from single- to multivariable treatment selection is also exemplified by a series of papers by Iniesta, Uher, and colleagues (Iniesta, Malki, et al., 2016; Iniesta et al., 2018; Uher et al., 2012). The authors used data from the Genome-based Therapeutic Drugs for Depression (GENDEP) study (Uher et al., 2009), in which participants were randomized to either a tricyclic antidepressant or a serotonin selective reuptake inhibitor. Initially, they tested the prognostic and prescriptive utility of each of nine variables is isolation (six symptom dimensions and three symptom cluster factors derived from those symptoms), and only found evidence for the anxiety symptom dimension as a moderator of antidepressant response (Uher et al., 2012). Recognizing the limitations of the single-variable approach, they then explored an expanded set of potential variables using a multivariable approach (Iniesta, Malki, et al., 2016). They found that models simultaneously including the effects of multiple variables predicted differential response to antidepressants with clinically meaningful accuracy, thus demonstrating the potential of multivariable approaches for treatment selection. Recently, they reanalyzed the GENDEP sample and combined genetic data with clinical and demographic variables to create drug-specific predictive models for antidepressant response that they validated in a held-out test sample (Iniesta et al., 2018).

Finally, other groups have used variants of the methods already described to address treatment selection questions (Cloitre, Petkova, Su, & Weiss, 2016; Westover et al., 2015). A recent review by Cohen and DeRubeis (2018) described over two dozen recent treatment selection efforts and revealed significant methodological heterogeneity, which can contribute to difficulties in detecting consistencies and inconsistencies in predictors, and creates a barrier to identifying "best practices" (Doove, Dusseldorp, Van Deun, & Van Mechelen, 2014).

Selection between Treatments Differing in Costs, Harms, and Other Dimensions

Treatments differ in terms of strength, cost, availability, and risk. While we have focused primarily on efficacy, it is but one consideration for treatment selection. Many health care systems follow a form of stratified care, the goal of which is to make best use of scarce treatment resources by organizing treatment options hierarchically. Briefer and less costly treatments are accessed first by many clients, so that more intensive options are available for those who are deemed to need them (Bower & Gilbody, 2005). Here, the more relevant question might be "What is the best way to allocate the stronger/costlier/less available/riskier (hereafter 'stronger') treatment?" Predictive modeling in a stratified care context should aim to enhance the efficient allocation of limited or costly resources, as well as to minimize patients' unnecessary exposure to treatments that require significant time commitments or are associated with increased side effect risk (Hingorani et al., 2013).

There are two ways in which the "stronger" treatments might produce superior group average change. One possibility is that individuals may vary in regard to the degree to which they benefit more from the stronger treatment versus the weaker one. In such cases, the identification of client characteristics that predict differential response between the stronger and the weaker interventions is of paramount importance. Alternatively, all clients might be expected to benefit more from the stronger treatment, and by similar amounts. In such cases, although allocation to the stronger treatment could not be based on differences in expected improvement, it could depend on prognosis in the weaker treatment. For example, the stronger treatment could be allocated to those with the worst prognoses in the weaker treatment. This would fit well within a stepped care model, based on the idea that those predicted to fare poorly in the weaker treatment would, eventually, be more likely to be given the stronger treatment as the next step.

Recently published efforts that use data from the National Health Service (NHS) Improving

Access to Psychological Therapy (IAPT) program highlight ways in which multivariable models may be used to guide stratified medicine in mental health. IAPT follows a hybrid stepped care/stratified care model, in which the majority of clients start with lower-intensity psychological interventions, and those who do not respond are stepped-up to higher intensity psychotherapy. Saunders, Cape, Fearon, and Pilling (2016) used latent profile analysis to create eight profiles that defined patient clusters, each of which described sets of baseline demographic data and symptom features that tended to co-occur. They successfully identified subsets of clients (those with profiles similar to each other) for whom outcomes were different in high-intensity treatment versus low-intensity psychological treatment. This model could be used to identify patients to send directly to the more intensive care, rather than having all patients start with the low-intensity treatment option. In a different sample, Delgadillo, Moreea, and Lutz (2016) created an index that generated predictions as to which clients were likely to achieve reliable and clinically significant reductions in depression or anxiety symptoms. Follow-up work using an index of case complexity yielded similar results in a separate sample of IAPT patients (Delgadillo et al., 2017). This case complexity index aimed to create treatment selection recommendations that clinicians can readily understand and interpret, and its utility is currently being tested in a clinical trial.

A relevant consideration in stepped care contexts is whether treatment selection decisions optimize patient outcomes, the efficient allocation of clinic resources, or a combination of these two concerns. A more intensive treatment might provide only marginal benefit for a patient, but at a high cost. From a population or clinic perspective, treatment selection decisions should weigh expected patient improvement versus expected cost. For example, a 5-point improvement may be worth a marginal cost of $5,000, but not $50,000.

Some work has examined the integration of expected harms into treatment selection decisions. For example, Kraemer and colleagues asked hypothetical patients to choose between profiles of expected improvements and side effects of two different drugs (Kraemer & Frank, 2010; Kraemer, Frank, & Kupfer, 2011). They were then able to quantify how patients valued the harms–benefit trade-off. Is 5 points of expected benefit worth twice as many side effects, or a 40% increased likelihood of a serious adverse event? Ultimately, this might be a personal decision for each patient. Information about patient harms must be presented to patients in informing treatment decisions.

In an analogous manner, a CT clinic that aims to increase its rate of success could use the findings from a prognostic study to inform the selection of patients for the clinic by taking preferentially those patients with high scores on a prognostic index. *But* the practical prediction question, from the patient's point of view, parallels the "placement" issue: "Which treatment is best for me?"

Future Directions

Evidence-based treatment selection in mental health today lags far behind where we need it to be. Few clinics or clinicians employ treatment selection algorithms, and few treatment selections algorithms have been robustly supported by multiple independent studies in independent samples. Clinicians need to be confident integrating these data and such approaches into their daily work. One conclusion from research to date is that single-variable models are unlikely to be valuable sources of information for strong recommendations about treatment selection for individual patients. Multivariable models represent a promising new direction, as we have argued, though validation studies are needed to establish the utility of emerging multivariable models.

Moreover, the pretreatment assessments that inform the treatment selection models of tomorrow will likely include biomarkers and other measures that promise to reveal prescriptive relationships, in addition to the self-report, environmental, demographic, and clinical variables that have been used in most treatment selection studies reported to date. Recent work has shown promise for neurobiological (Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015; Jollans & Whelan, 2016; Pizzagalli, 2011; Stephan et al., 2017), and neurocognitive and behavioral variables (Webb et al., 2018), as well as measures of immune function (Uher et al., 2014). Several recent studies have been designed specifically to generate knowledge relevant to outcome prediction in depression treatment (Dunlop et al., 2012; Green et al., 2017; Grieve et al., 2013; Lam et al., 2016; Trivedi et al., 2016; Williams, 2017). They feature potential biomarkers, including

information from neuroimaging (McGrath et al., 2013; Pizzagalli et al., 2018) and genetic tests (Iniesta et al., 2018; Lam et al., 2016; Ward et al., 2018).

Research aiming to identify biological predictors of mental health outcomes is still in its infancy. Although many promising findings have been reported (Drysdale et al., 2017), very few of these have been replicated in independent samples (Woo, Chang, Lindquist, & Wager, 2017). Additionally, biological measures can be costly and difficult to collect in routine clinical settings. As our understanding and awareness of these variables increases, it will be important to demonstrate the added value of more expensive predictors, or to link them to more easily assessible variables that can serve as "proxies" until such times as the cost of measurement decreases.

Moreover, for treatment selection to be effective, clinicians and clients must have available different treatment options from which to choose. For example, treatment selection at the level of intervention "packages" (i.e., ADM vs. CBT vs. IPT) would not be useful in some rural locations where evidence-based psychotherapy is unavailable, and the majority of mental health treatment is delivered by family medicine doctors who prescribe antidepressants. Similarly, a psychotherapist who is trained only to provide CBT might not be open to a recommendation that his or her client would be better suited for IPT. In the United States, the Veterans Administration health care system is an excellent candidate for treatment selection, as it has established an infrastructure for training and delivery of a variety of evidence-based treatments for PTSD, including CPT, PE, PCT, and eye movement desensitization and reprocessing (EMDR).

Another context ripe for treatment selection is the NHS IAPT system in the United Kingdom, which treats over 560,000 patients per year, collects a standard set of baseline predictive variables, and offers different forms of psychotherapy at different levels of intensity (Clark, 2018). Both stratified medicine (determining for whom low-intensity treatment is sufficient and who should begin with high-intensity treatment) and treatment selection between equivalent treatments (selecting which psychotherapy among a set of equally effective interventions would be best for an individual) would be possible within IAPT. We are currently running a prediction tournament in which 13 teams have been given a large sample of anonymized patient data that include the set of universally collected baseline variables and treatment outcomes. If the models that are developed prove useful, they could be instantiated in IAPT clinics across the United Kingdom and could provide individualized outcome predictions that could be used by clinicians and clients in a shared decision-making process to improve the way in which treatments are allocated. The potential impact that treatment selection could have at this scale should not be underestimated. If model-informed treatment allocation could improve IAPT's current 50% recovery rate by even 5%, it could result in 28,000 more individuals recovering each year.

Conclusions

Moving beyond "one size fits all" (the title of this chapter) has been a goal of both clinicians and researchers for many decades. Recognizing that unique individuals will respond differently to treatment, clinicians have long attempted to personalize or adapt treatments to their clients. Recent successes in precision medicine in other areas have inspired new research efforts in clinical psychology.

Although we have not yet addressed with confidence the question that Paul asked so many decades ago—what works for whom?—we are getting closer to a time when research will inform the questions that clinicians and clients are asking. We have recognized that efforts to guide treatment selection based on a single feature of the client are misguided. Despite the appealing simplicity of such studies, they have had limited impact on client care (Simon & Perlis, 2010). There is great value for clinicians in knowing when and why to be skeptical of the research literature, and simple single-variable prediction results should invite such caution. Similarly, many of the more complex multivariate approaches used today are limited by the exploratory nature of the models. Again, clinicians and clients are wise to demand that this work continue, so that promising findings are put to the test of validation and replication. Despite these cautions, the PAI and many of the other multivariable treatment selection approaches we have reviewed are on the cutting edge of efforts to integrate science and practice.

This knowledge matters. Clinicians want to provide individuals struggling with mental health problems with treatments that work. Reducing the number of ineffective treatments to

which individuals are exposed will reduce their suffering and will benefit communities through reducing the loss of work productivity associated with mental illness (Layard, Clark, Knapp, & Mayraz, 2007). We all will benefit from an improved understanding of what will work for whom.

References

Agniel, D., Kohane, I. S., & Weber, G. M. (2018). Biases in electronic health record data due to processes within the healthcare system: Retrospective observational study. *British Medical Journal, 361,* k1479.

American Psychiatric Association. (2010). *Practice guideline for the treatment of patients with major depressive disorder.* Arlington, VA: Author.

American Psychological Association. (2006). Evidence-based practice in psychology: APA presidential task force on evidence-based practice. *American Psychologist, 61*(4), 271–285.

Amsterdam, J. D., Lorenzo-Luaces, L., & DeRubeis, R. J. (2016). Step-wise loss of antidepressant effectiveness with repeated antidepressant trials in bipolar II depression. *Bipolar Disorders, 18,* 563–570.

Amsterdam, J. D., & Shults, J. (2009). Does tachyphylaxis occur after repeated antidepressant exposure in patients with bipolar II major depressive episode? *Journal of Affective Disorders, 115,* 234–240.

Amsterdam, J. D., Williams, D., Michelson, D., Adler, L. A., Dunner, D. L., et al. (2009). Tachyphylaxis after repeated antidepressant drug exposure in patients with recurrent major depressive disorder. *Neuropsychobiology, 59,* 227–233.

Ashar, Y. K., Chang, L. J., & Wager, T. D. (2017). Brain mechanisms of the placebo effect: An affective appraisal account. *Annual Review of Clinical Psychology, 13,* 73–98.

Ashley, E. A. (2016). Towards precision medicine. *Nature Reviews Genetics, 17,* 507–522.

Barber, J. P., & Muenz, L. R. (1996). The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: Empirical findings from the Treatment for Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64,* 951–958.

Barbui, C., Cipriani, A., Patel, V., Ayuso-Mateos, J. L., & van Ommeren, M. (2011). Efficacy of antidepressants and benzodiazepines in minor depression: Systematic review and meta-analysis. *British Journal of Psychiatry, 198,* 11–16.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press

Blau, G., & DiMino, J. (2018). Impact of brief counseling on nonurgent prior- versus never-counseled samples. *Journal of College Student Psychotherapy.* [Epub ahead of print]

Boswell, J. F., McAleavey, A. A., Castonguay, L. G., Hayes, J. A., & Locke, B. D. (2012). Previous mental health service utilization and change in clients' depressive symptoms. *Journal of Counseling Psychology, 59,* 368–378.

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. *British Journal of Psychiatry, 186,* 11–17.

Byrne, S. E., & Rothschild, A. J. (1998). Loss of antidepressant efficacy during maintenance therapy: Possible mechanisms and treatments. *Journal of Clinical Psychiatry, 59,* 279–288.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7–18.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., & Trivedi, M. H. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *Lancet Psychiatry, 3,* 243–250.

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT Program. *Annual Review of Clinical Psychology, 14,* 159–183.

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. *British Journal of Psychiatry Open, 2,* 101–106.

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology, 14,* 209–236.

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J., & Driessen, E. (2019). A demonstration of a multimethod variable selection approach for treatment selection: Recommending cognitive–behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychotherapy Research*. [Epub ahead of print]

Connolly, K. R., & Thase, M. E. (2011). The clinical management of bipolar disorder: A review of evidence-based guidelines. *Primary Care Companion to CNS Disorders, 13,* 10r01097.

Cook, J. M., Dinnen, S., Simiola, V., Thompson, R., & Schnurr, P. P. (2014). VA residential provider perceptions of dissuading factors to the use of two evidence-based PTSD treatments. *Professional Psychology: Research and Practice, 45,* 136–142.

Cook, J. M., Simiola, V., Hamblen, J. L., Bernardy, N., & Schnurr, P. P. (2017). The influence of patient readiness on implementation of evidence-based PTSD treatments in Veterans Affairs residential programs. *Psychological Trauma: Theory, Research, Practice, and Policy, 9,* 51–58.

Craske, M. G., Meuret, A. E., Ritz, T., Treanor, M., & Dour, H. J. (2016). Treatment for anhedonia: A neuroscience driven approach. *Depression and Anxiety, 33,* 927–938.

Cuijpers, P., Huibers M. J., & Furukawa, T. A. (2017). The need for research on treatments of chronic depression. *JAMA Psychiatry, 74,* 242–243.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582.

Dawes, R. M. (2005). The ethical implications of Paul Meehl's work on comparing clinical versus actuarial prediction methods. *Journal of Clinical Psychology, 61,* 1245–1255.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1674.

Deisenhofer, A. K., Delgadillo, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., et al. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety, 35,* 541–550.

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

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy, 79,* 15–22.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating research on prediction into individualized treatment recommendations: A demonstration. *PLOS ONE, 9,* e83875.

DeRubeis, R. J., Gelfand, L. A., German, R. E., Fournier, J. C., & Forand, N. R. (2014). Understanding processes of change: How some patients reveal more than others—and some groups of therapists less—about what matters in psychotherapy. *Psychotherapy Research, 24,* 419–428.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., et al. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62,* 409–416.

Doove, L. L., 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. *Advances in Data Analysis and Classification, 8,* 403–425.

Driessen, E., Cuijpers, P., Hollon, S. D., & Dekker, J. J. (2010). Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression?: A meta-analysis. *Journal of Consulting and Clinical Psychology, 78,* 668–680.

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., et al. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine, 23,* 28–38.

Dunlop, B. W., Kelley, M. E., Aponte-Rivera, V., Mletzko-Crowe, T., Kinkead, B., Ritchie, J. C., et al. (2017). Effects of patient preferences on outcomes in the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study. *American Journal of Psychiatry, 174,* 546–556.

Dunlop, B. W., Kelley, M. E., Mletzko, T. C., Velasquez, C. M., Craighead, W. E., & Mayberg, H. S. (2012). Depression beliefs, treatment preference, and outcomes in a randomized trial for major depressive disorder. *Journal of Psychiatric Research, 46,* 375–381.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., et al. (1989). National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. *Archives of General Psychiatry, 46,* 971–982.

Fineberg, N. A., Brown, A., Reghunandanan, S., & Pampaloni, I. (2012). Evidence-based pharmacotherapy of obsessive–compulsive disorder. *International Journal of Neuropsychopharmacology, 15,* 1173–1191.

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., et al. (2010). Antidepressant drug effects and depression severity: A patient-level meta-analysis. *Journal of the American Medical Association, 303,* 47–53.

Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Gallop, R., Amsterdam, J. D., & Hollon, S. D. (2008). Antidepressant medications v. cognitive therapy in people with depression with or without personality disorder. *British Journal of Psychiatry, 192,* 124–129.

Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Hollon, S. D., Amsterdam, J. D., & Gallop, R. (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *Journal of Consulting and Clinical Psychology, 77,* 775–787.

Furukawa, T. A., Weitz, E. S., Tanaka, S., Hollon, S. D., Hofmann, S. G., Andersson, G., et al. (2017). Initial severity of depression and efficacy of cognitive-behavioural therapy: Individual-participant data meta-analysis of pill-placebo-controlled trials. *British Journal of Psychiatry, 210,* 190–196.

Gabrieli, J. D., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron, 85,* 11–26.

Gillan, C. M., & Daw, N. D. (2016). Taking psychiatry research online. *Neuron, 91,* 19–23.

Gillan, C. M., & Whelan, R. (2017). What big data can do for treatment in psychiatry. *Current Opinion in Behavioral Sciences, 18,* 34–42.

Green, E., Goldstein-Piekarski, A. N., Schatzberg, A. F., Rush, A. J., Ma, J., & Williams, L. (2017). Personalizing antidepressant choice by sex, body mass

index, and symptom profile: An iSPOT-D report. *Personalized Medicine in Psychiatry, 1,* 65–73.

Grenyer, B. F., Deane, F. P., & Lewis, K. L. (2008). Treatment history and its relationship to outcome in psychotherapy for depression. *Counselling and Psychotherapy Research, 8,* 21–27.

Grieve, S. M., Korgaonkar, M. S., Etkin, A., Harris, A., Koslow, S. H., Wisniewski, S., et al. (2013). Brain imaging predictors and the international study to predict optimized treatment for depression: Study protocol for a randomized controlled trial. *Trials, 14,* 224.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law, 2,* 293–323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12,* 19–30.

Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine, 363,* 301–304.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61,* 155–163.

Hingorani, A. D., van der Windt, D. A., Riley, R. D., Abrams, K., Moons, K. G., Steyerberg, E. W., et al. (2013). Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *British Medical Journal, 346,* e5793.

Hirschfeld, R. M. (2001). The comorbidity of major depression and anxiety disorders: Recognition and management in primary care. *Primary Care Companion to the Journal of Clinical Psychiatry, 3,* 244–254.

Holmes, E. A., Ghaderi, A., Harmer, C. J., Ramchandani, P. G., Cuijpers, P., Morrison, A. P., et al. (2018). The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *Lancet Psychiatry, 5,* 237–286.

Huibers, M. J., Cohen, Z. D., Lemmens, L. H., Arntz, A., Peeters, F. P., Cuijpers, P., 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.

Iniesta, R., Hodgson, K., Stahl, D., Malki, K., Maier, W., Rietschel, M., et al. (2018). Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables. *Scientific Reports, 8,* Article No. 5530.

Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., et al. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric Research, 78,* 94–102.

Iniesta, R., Stahl, D., & McGuffin, P. (2016). Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological Medicine, 46,* 2455–2465.

Janes, H., Pepe, M. S., Bossuyt, P. M., & Barlow, W. E. (2011). Measuring the performance of markers for guiding treatment decisions. *Annals of Internal Medicine, 154*(4), 253–259.

Jollans, L., & Whelan, R. (2016). The clinical added value of imaging: A perspective from outcome prediction. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1,* 423–432.

Katsnelson, A. (2013). Momentum grows to make "personalized" medicine more "precise." *Nature Medicine, 19,* 249.

Keefe, J. R., Wiltsey-Stirman, S., Cohen, Z. D., DeRubeis, R. J., Smith, B. N., & Resick, P. (2018). In rape-trauma PTSD, patient characteristics indicate which trauma-focused treatment they are most likely to complete. *Depression and Anxiety, 35,* 330–338.

Kessler, R. C. (2018). The potential of predictive analytics to provide clinical decision support in depression treatment planning. *Current Opinion in Psychiatry, 31,* 32–39.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62,* 617–627.

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., et al. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences, 26,* 22–36.

Khan, A., Leventhal, R. M., Khan, S. R., & Brown, W. A. (2002). Severity of depression and response to antidepressants and placebo: An analysis of the Food and Drug Administration database. *Journal of Clinical Psychopharmacology, 22,* 40–45.

Kircanski, K., LeMoult, J., Ordaz, S., & Gotlib, I. H. (2017). Investigating the nature of co-occurring depression and anxiety: Comparing diagnostic and dimensional research approaches. *Journal of Affective Disorders, 216,* 123–135.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine, 5,* e45.

Kocsis, J. H., Leon, A. C., Markowitz, J. C., Manber, R., Arnow, B., Klein, D. N., 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. *Journal of Clinical Psychiatry, 70,* 354–361.

Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P, Wobrock, T., et al. (2016). Multisite prediction of 4-week and 52-week treatment outcomes

in patients with first-episode psychosis: A machine learning approach. *Lancet Psychiatry, 3,* 935–946.

Kraemer, H. C. (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: A parametric approach. *Statistics in Medicine, 32,* 1964–1973.

Kraemer, H. C., & Frank, E. (2010). Evaluation of comparative treatment trials: Assessing clinical benefits and risks for patients, rather than statistical effects on measures. *Journal of the American Medical Association, 304,* 683–684.

Kraemer, H. C., Frank, E., & Kupfer, D. J. (2011). How to assess the clinical impact of treatments on patients, rather than the statistical impact of treatments on measures. *International Journal of Methods in Psychiatric Research, 20,* 63–72.

Lam, R. W., Milev, R., Rotzinger, S., Andreazza, A. C., Blier, P., Brenner, C., 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. *National Institute Economic Review, 202,* 90–98.

Leykin, Y., Amsterdam, J. D., DeRubeis, R. J., Gallop, R., Shelton, R. C., & Hollon, S. D. (2007). Progressive resistance to a selective serotonin reuptake inhibitor but not to cognitive therapy in the treatment of major depression. *Journal of Consulting and Clinical Psychology, 75,* 267–276.

Leykin, Y., DeRubeis, R. J., Gallop, R., Amsterdam, J. D., Shelton, R. C., & Hollon, S. D. (2007). The relation of patients' treatment preferences to outcome in a randomized clinical trial. *Behavior Therapy, 38,* 209–217.

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology, 50,* 7–36.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science, 4,* 390–398.

Linardon, J., Wade, T. D., de la Piedad Garcia, X., & Brennan, L. (2017). The efficacy of cognitive-behavioral therapy for eating disorders: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 85,* 1080–1094.

Lo, A., Chernoff, H., Zheng, T., & Lo, S.-H. (2015). Why significant variables aren't automatically good predictors. *Proceedings of the National Academy of Sciences of the USA, 112,* 13892–13897.

Lorenzo-Luaces, L., DeRubeis, R. J., & Bennett, I. M. (2015). Primary care physicians' selection of low-intensity treatments for patients with depression. *Family Medicine, 47,* 511–516.

Lorenzo-Luaces, L., DeRubeis, R. J., van Straten, A., & Tiemens, B. (2017). A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. *Journal of Affective Disorders, 213,* 78–85.

Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M., Mulder, R. T., et al. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry, 190,* 496–502.

Lutz, W., Hofmann, S. G., Rubel, J., Boswell J. F., Shear, M. K., Gorman, J. M., et al. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology, 82,* 287–297.

Lutz, W., Saunders, S. M., Leon, S. C., Martinovich, Z., Kosfelder, J., Schulte, D., et al. (2006). Empirically and clinically useful decision making in psychotherapy: Differential predictions with treatment response models. *Psychological Assessment, 18,* 133–141.

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., et al. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron, 45,* 651–660.

McCullough, J. P., Jr. (2003). *Treatment for chronic depression: Cognitive behavioral analysis system of psychotherapy (CBASP).* New York: Guilford Press.

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. *Psychological Medicine, 45,* 693–704.

McGrath, C. L., Kelley, M. E., Holtzheimer, P. E., Dunlop, B. W., Craighead, W. E., Franco, A. R., et al. (2013). Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry, 70,* 821–829.

McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013). Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: A meta-analytic review. *Journal of Clinical Psychiatry, 74,* 595–602.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46,* 806–834.

Mergl, R., Henkel, V., Allgaier, A. K., Kramer, D., Hautzinger, M., Kohnen, R., 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. *Psychotherapy and Psychosomatics, 80,* 39–47.

Mickey, R. M., & Greenland, S. (1989). The impact of confounder selection criteria on effect estimation. *American Journal of Epidemiology, 129,* 125–137.

National Health Service. (2016). *Psychological therapies, annual report on the use of IAPT services—England 2015–16.* London: Health and Social Care information Centre.

National Institute for Health and Clinical Excellence. (2009). *Depression: Treatment and management of depression in adults.* London: Author.

National Research Council. (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease.* Washington, DC: National Academies Press.

Nemeroff, C. B., Heim, C. M., Thase, M. E., Klein, D. N., Rush, A. J., Schatzberg, A. F., et al. (2003). Differential responses to psychotherapy versus pharmacotherapy in patients with chronic forms of major depression and childhood trauma. *Proceedings of the National Academy of Sciences of the USA, 100,* 14293–14296.

Niles, A. N., Loerinc, A. G., Krull, J. L., Roy-Byrne, P., Sullivan, G., Sherbourne, C. D., et al. (2017). Advancing personalized medicine: Application of a novel statistical method to identify treatment moderators in the coordinated anxiety learning and management study. *Behavior Therapy, 48,* 490–500.

Niles, A. N., Wolitzky-Taylor, K. B., Arch, J. J., & Craske, M. G. (2017). Applying a novel statistical method to advance the personalized treatment of anxiety disorders: A composite moderator of comparative drop-out from CBT and ACT. *Behaviour Research and Therapy, 91,* 13–23.

Nuzzo, R. (2014). Statistical errors. *Nature, 506,* 150–152.

Osei-Bonsu, P. E., Bolton, R. E., Stirman, S. W., Eisen, S. V., Herz, L., & Pellowe, M. E. (2017). Mental health providers' decision-making around the implementation of evidence-based treatment for PTSD. *Journal of Behavioral Health Services and Research, 44,* 213–223.

Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., et al. (2004). EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science, 304,* 1497–1500.

Papakostas, G. I., & Fava, M. (2010). *Pharmacotherapy for depression and treatment-resistant depression.* Singapore: World Scientific.

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

Pauker, S. G., & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *New England Journal of Medicine, 302,* 1109–1117.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 31,* 109–118.

Perlis, R. H. (2013). A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biological Psychiatry, 74,* 7–14.

Perlis, R. H. (2016). Abandoning personalization to get to precision in the pharmacotherapy of depression. *World Psychiatry, 15,* 228–235.

Perlis, R. H., Fijal, B., Dharia, S., Heinloth, A. N., & Houston, J. P. (2010). Failure to replicate genetic associations with antidepressant treatment response in duloxetine-treated patients. *Biological Psychiatry, 67,* 1110–1113.

Perlis, R., Iosifescu, D., Castro, V., Murphy, S., Gainer, V., Minnier, J., et al. (2012). Using electronic medical records to enable large-scale studies in psychiatry: Treatment resistant depression as a model. *Psychological Medicine, 42,* 41–50.

Perlis, R. H., Patrick, A., Smoller, J. W., & Wang, P. S. (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,* 2227–2236.

Pizzagalli, D. A. (2011). Frontocingulate dysfunction in depression: Toward biomarkers of treatment response. *Neuropsychopharmacology, 36,* 183–206.

Pizzagalli, D. A., Webb, C. A., Dillon, D. G., Tenke, C. E., Kayser, J., Goer, F., et al. (2018). Pretreatment rostral anterior cingulate cortex theta activity in relation to symptom improvement in depression: A randomized clinical trial. *JAMA Psychiatry, 75,* 547–554.

Preference Collaborative Review Group. (2008). Patients' preferences within randomised trials: Systematic review and patient level meta-analysis. *British Medical Journal, 337,* a1864.

Raza, G. T., & Holohan, D. R. (2015). Clinical treatment selection for posttraumatic stress disorder: Suggestions for researchers and clinical trainers. *Psychological Trauma: Theory, Research, and Practice, 7,* 547–554.

Renjilian, D. A., Perri, M. G., Nezu, A. M., McKelvey, W. F., Shermer, R. L., & Anton, S. D. (2001). Individual versus group therapy for obesity: Effects of matching participants to their treatment preferences. *Journal of Consulting and Clinical Psychology, 69,* 717–721.

Rosell, R., Carcereny, E., Gervais, R., Vergnenegre, A., Massuti, B., Felip, E., 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 Oncology, 13,* 239–246.

Rosen, C. S., Matthieu, M., Stirman, S. W., Cook, J., Landes, S., Bernardy, N. C., et al. (2016). A review of studies on the system-wide implementation of evidence-based psychotherapies for posttraumatic stress disorder in the Veterans Health Administration. *Administration and Policy in Mental Health and Mental Health Services Research, 43,* 957–977.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276–1284.

Saunders, R., Cape, J., Fearon, P., & Pilling, S. (2016). Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. *Journal of Affective Disorders, 197,* 107–115.

Schleidgen, S., Klingler, C., Bertram, T., Rogowski, W. H., & Marckmann, G. (2013). What is personalized medicine: Sharpening a vague term based on a systematic literature review. *BMC Medical Ethics, 14,* 55.

Schweizer, S., Cohen, Z., Hayes, R., DeRubeis, R., Crane, C., Kuyken, W., et al. (2019). *Relapse prevention for antidepressant medication (ADM) responders with recurrent depression: Deciding between maintenance ADM and mindfulness-based cognitive therapy.* Manuscript submitted for publication.

Simon, G. E., Johnson, E., Lawrence, J. M., Rossom, R. C., Ahmedani, B., Lynch, F. L., et al. (2018). Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *American Journal of Psychiatry, 175,* 951–960.

Simon, G. E., & Perlis, R. H. (2010). Personalized medicine for depression: Can we match patients with treatments? *American Journal of Psychiatry, 167,* 1445–1455.

Smagula, S. F., Wallace, M. L., Anderson, S. J., Karp, J. F., Lenze, E. J., Mulsant, B. H., 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. *Journal of Psychiatric Research, 81,* 112–118.

Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2009). The meta-analysis of clinical judgment project: Effects of experience on judgment accuracy. *Counseling Psychologist, 37,* 350–399.

Stephan, K. E., Schlagenhauf, F., Huys, Q. J., Raman, S., Aponte, E. A., Brodersen, K. H., et al. (2017). Computational neuroimaging strategies for single patient predictions. *NeuroImage, 145,* 180–199.

Swift, J. K., & Callahan, J. L. (2009). The impact of client treatment preferences on outcome: A meta-analysis. *Journal of Clinical Psychology, 65,* 368–381.

Swift, J. K., Callahan, J. L., & Vollmer, B. M. (2011). Preferences. *Journal of Clinical Psychology, 67,* 155–165.

Trivedi, M. H., McGrath, P. J., Fava, M., Parsey, R. V., Kurian, B. T., Phillips, M. L., et al. (2016). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *Journal of Psychiatric Research, 78,* 11–23.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Uher, R., Huezo-Diaz, P., Perroud, N., Smith, R., Rietschel, M., Mors, O., et al. (2009). Genetic predictors of response to antidepressants in the GENDEP project. *Pharmacogenomics Journal, 9,* 225–233.

Uher, R,, Perlis, R., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., et al. (2012). Depression symptom dimensions as predictors of antidepressant treatment outcome: Replicable evidence for interest-activity symptoms. *Psychological Medicine, 42,* 967–980.

Uher, R., Tansey, K. E., Dew, T., Maier, W., Mors, O., Hauser, J., et al. (2014). An inflammatory biomarker as a differential predictor of outcome of depression treatment with escitalopram and nortriptyline. *American Journal of Psychiatry, 171,* 1278–1286.

van Straten, A., Tiemens, B., Hakkaart, L., Nolen, W., & Donker, M. (2006). Stepped care vs. matched care for mood and anxiety disorders: A randomized trial in routine practice. *Acta Psychiatrica Scandinavica, 113,* 468–476.

Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (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 Research, 253,* 174–181.

Vittengl, J. R., Jarrett, R. B., Weitz, E., Hollon, S. D., Twisk, J., Cristea, I., et al. (2016). Divergent outcomes in cognitive-behavioral therapy and pharmacotherapy for adult depression. *American Journal of Psychiatry, 173,* 481–490.

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110,* 639–644.

Wallace, M. L., Frank, E., & Kraemer, H. C. (2013). A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *JAMA Psychiatry, 70,* 1241–1247.

Ward, J., Graham, N., Strawbridge, R., Ferguson, A., Jenkins, G., Chen, W., et al. (2018). Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: Meta-analysis of three treatment cohorts. *PLOS ONE, 13*(9), e0203896.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician, 70,* 129–133.

Webb, C. A., Trivedi, M. H., Cohen, Z. D., Dillon, D. G., Fournier, J. C., Goer, F., et al. (2018). Personalized prediction of antidepressant v. placebo response: Evidence from the EMBARC study. *Psychological Medicine.* [Epub ahead of print]

Weitz, E. S., Hollon, S. D., Twisk, J., van Straten, A., Huibers, M. J., David, D., et al. (2015). Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: An individual patient data meta-analysis. *JAMA Psychiatry, 72,* 1102–1109.

Westover, A. N., Kashner, T. M., Winhusen, T. M., Golden, R. M., Nakonezny, P. A., Adinoff, B., et al. (2015). A systematic approach to subgroup analyses

in a smoking cessation trial. *American Journal of Drug and Alcohol Abuse, 41,* 498–507.

Williams, L. M. (2017). Getting personalized: Brain scan biomarkers for guiding depression interventions. *American Journal of Psychiatry, 174,* 503–505.

Wiltsey Stirman, S., Cohen, Z., Lunney, C., DeRubeis, R., Wiley, J., & Schnurr, P. (2019). *A prognostic index to inform selection of a trauma-focused or non-trauma-focused treatment for PTSD.* Manuscript submitted for publication.

Winter, S. E., & Barber, J. P. (2013). Should treatment for depression be based more on patient preference? *Patient Preference and Adherence, 7,* 1047–1057.

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience, 20,* 365–377.

World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates.* Geneva: Author.

Zilcha-Mano, S., Keefe, J. R., Chui, H., Rubin, A., Barrett, M. S., & Barber, J. P. (2016). Reducing dropout in treatment for depression: Translating dropout predictors into individualized treatment recommendations. *Journal of Clinical Psychiatry, 77,* e1584–e1590.