Chapter 19: Personalized Treatment Approaches

Zachary D. Cohen, Jaime Delgadillo, Robert J. DeRubeis

Abbreviations

3DP	three dimensions of personalization
ACT	acceptance and commitment therapy
ATI	aptitude-by-treatment interactions
BASIC ID	behavior, affect, sensation, imagery, cognition, interpersonal relationships, and drug/biology
CBASP	cognitive behavioral analysis system of psychotherapy
CBT	cognitive behavioral therapy
CPT	cognitive-processing therapy
DBT	dialectic behavioral therapy
EMA	ecological momentary assessment
EMR	electronic medical records
IPT	interpersonal therapy
LASSO	least absolute shrinkage and selection operator
MBCT	mindfulness-based cognitive therapy
MDD	major depressive disorder
NCCMH	National Collaborating Centre for Mental Health
OCD	obsessive-compulsive disorder
PAI	personalized advantage index
PATH	predictive approaches to treatment effect heterogeneity
PE	prolonged exposure
PROGRESS	PROGnosis RESearch Strategy
PTSD	post-traumatic stress disorder
RCT	randomized clinical trial
SMART	sequential multiple assignment randomized trials

STRATOS	STRengthening Analytical Thinking for Observational Studies
TAU	treatment as usual
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis
WEIRD	western, educated, industrialized, rich and democratic

OUTLINE

Introduction: The Personalized Treatment Imperative
A Conceptual Framework for Personalized Psychotherapy
Intuitive and Theoretical Approaches to Psychotherapy Personalization
- Intuitive Approaches to Psychotherapy Personalization
- Theoretical Approaches to Psychotherapy Personalization
- Summary: The Road to Precision Starts With Trial and Error

Contemporary Applications of Personalization in Psychotherapy
- Selecting the Intensity of Psychotherapeutic Interventions in Stratified Care
- Selecting Treatment Packages
- Selecting Treatment Components or Techniques
- Adjusting Psychotherapy as It Unfolds

Methodological and Statistical Issues for Personalized Psychotherapy
- Evaluation
- Constructing Data-driven Personalized Treatment Decision Support Systems

Challenges: Analytic, Methodological, and Statistical Problems with the Existing Literature on Personalized Psychotherapy
- Defining and Reporting Personalization Systems
- Current Limitations of Machine Learning
- Evaluation

Conclusions
- An Evolutionary Perspective
- Current Challenges and Future Directions

Abstract

In the modern history of psychotherapy, understanding the individual patient and how to optimize treatment for each individual has been an important challenge. For the therapist, personalization often has meant deciding which treatments to offer based on clinical assessment and formulation, or deciding moment-to-moment what techniques to employ, given what has been happening in therapy with the patient, informed by a particular theoretical framework, or based on clinical judgment. Efforts to improve outcomes through treatment personalization based on expert clinical assessments and psychological theory have been largely unfruitful, despite their intuitive appeal. Furthermore, numerous studies over 50 years have highlighted the limitations of clinical judgment, making a compelling case for the use of data-driven, actuarial methods to enhance

treatment decisions. New statistical approaches, including machine learning, have propelled advancements in our ability to develop data-derived decision aids that can be tested and easily disseminated. Personalization can take several forms, including determining the appropriate level of care or treatment package, selecting and ordering of treatment components or modules, identifying therapeutic targets, or matching patients to specific therapists. This chapter reviews the evidence for different forms of treatment personalization, from intuitive and theory-based approaches to contemporary data-driven approaches. After summarizing the limitations of the current field, the chapter concludes with a discussion of what the future could hold.

Introduction: The Personalized Treatment Imperative

People respond differently to psychotherapy, even if they present with similar problems and are treated by the same therapist with similar methods (Kiesler, 1966). This phenomenon is often called the *heterogeneity of treatment effect* (Varadhan et al., 2013). Empirical studies highlighting the prognostic value of individual differences were conducted early in the 20th century (e.g., Strupp, 1964), spurring many psychotherapists to argue for the need to personalize and adjust interventions with regard to features other than diagnosis (Garfield, 1996). This view has been ubiquitous in clinical psychology for many decades, and has its corollary in Gordon Paul's classic question to the field: "*What* treatment, by *whom*, is most effective for *this* individual with *that* specific problem, and under *which* set of circumstances?" (Paul, 1967, p. 111). Accordingly, psychotherapists have been attempting to deliver personalized therapy for decades, but there is tremendous diversity in the approaches they have used. Recent efforts to personalize treatment across mental health have focused on selecting the optimal psychotherapy for an individual among the available evidence-based approaches (Cohen & DeRubeis, 2018). In this chapter, we present a review of several approaches to personalization and propose conceptual frameworks aimed at helping readers [1] to understand, [2] to categorize, and [3] to appraise the methodological rigor of these different approaches. Strengths and weaknesses of different approaches will be illustrated using examples from the literature. Overall, this chapter aims to survey the landscape of personalization from a historical and methodological perspective, clarifying key concepts, and concluding with a consideration of future research directions that promise to enhance the precision and effectiveness of personalization.

This is an auspicious moment in the evolution of personalization in psychotherapy. The increasing emphasis on evidence-based psychological treatments, combined with rapid developments in statistical methodologies and the availability of datasets that can take advantage of these developments, are driving a proliferation of research programs that are giving rise to a new field of precision mental health care. As investigators begin to implement more complex methodologies, they discover a multitude of additional hurdles that must be surmounted in order for their contributions to be ready to be used by the clinicians and patients who might benefit from them. In this chapter, we will examine a variety of questions and methods that come into play in research on, and in the implementation of, personalized treatment approaches. Along the way we propose a lexicon that we hope will organize and clarify the many conceptual and methodological distinctions that abound in this work, and we will present examples of research and implementation efforts that can serve as springboards for programs going forward.

The chapter is organized in six parts. The second section introduces a framework that comprehensively maps out the various approaches to personalization seen in the psychotherapy literature. This framework offers a taxonomy of distinctive dimensions of personalization and serves to organize a discussion of the variety of traditions and methods in this area. In the third section, we describe historical efforts to characterize or systematize the personalization of psychotherapy. The fourth section provides a description of contemporary innovations, focusing on data-driven algorithms, and highlighting promising lines of research and cutting-edge demonstrations. In the fifth section we discuss relevant methodological and statistical issues. In the sixth section, we describe key challenges that must be addressed if research on personalization is to play a major role in how psychotherapy is practiced. We end with suggestions for how these barriers can be overcome, and a discussion of future directions.

A Conceptual Framework for Personalized Psychotherapy

The framework in Figure 19.1 depicts, as dimensions, fundamental ways in which approaches to personalization differ from each other. This is proposed as an organizing framework to understand and catalog the variety of ways in which personalized therapy has been pursued and promoted. The framework, called the *Three Dimensions of Personalization (3DP),* enables us to describe the types of decisions that can be made to personalize treatment, the timing of these decisions, and the methods used to inform such decisions. Multiple combinations of features across these three dimensions are possible, thus enabling psychotherapy

researchers to derive a highly specific taxonomy of features that characterizes current approaches to personalization. In what follows, we will unpack these concepts and offer some illustrative examples.

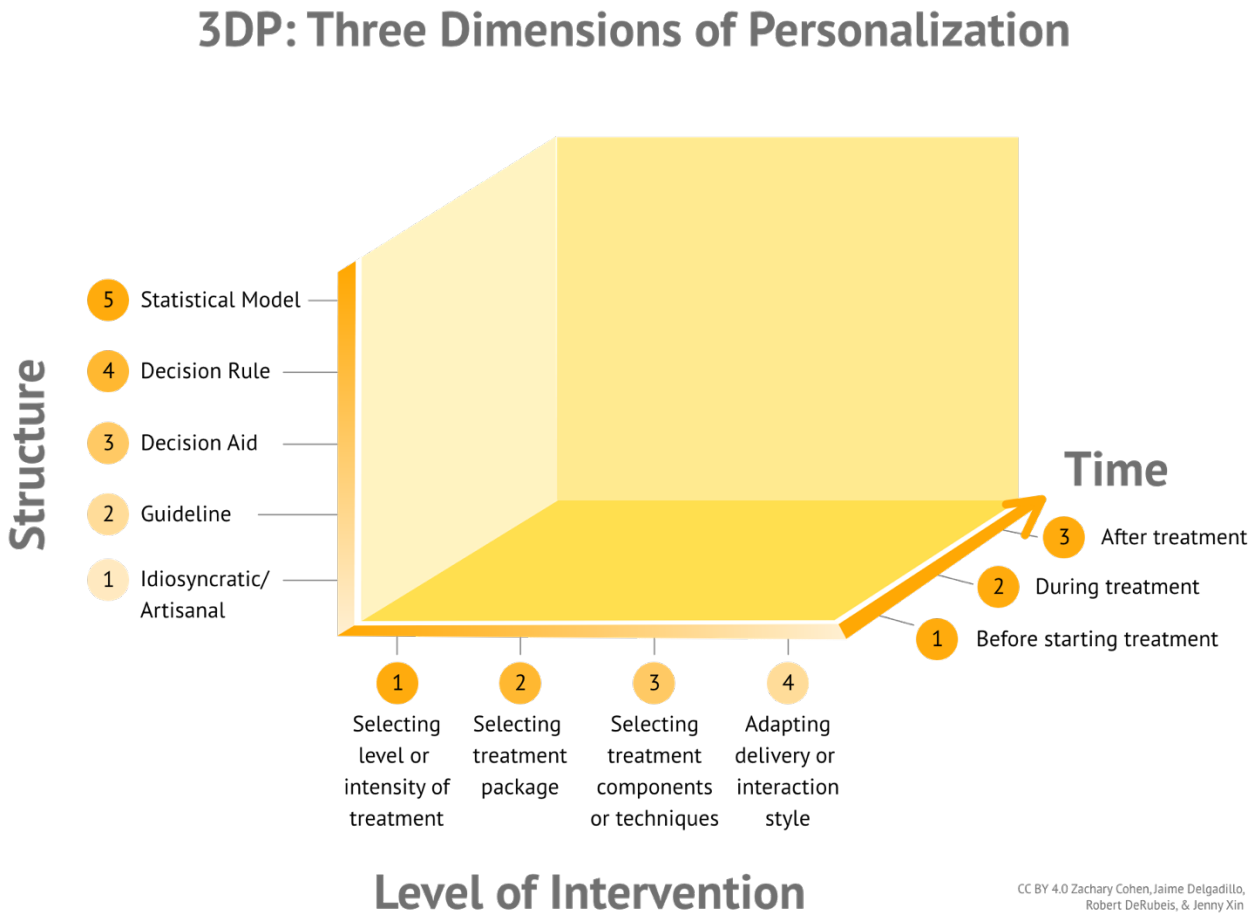

Figure 19.1. Three Dimensions of Personalization (3DP)

The *Time* dimension characterizes the stage at which personalization decisions are made along a patient's treatment pathway. This could be before the start of treatment (Time 1 or 3DP:T1), across the course of treatment (3DP:T2), or after treatment has ended (3DP:T3) (e.g., relapse prevention). The *Level of Intervention* dimension describes the level of specificity of treatment decisions, pertaining to macro-level decisions (e.g., high intensity versus low intensity interventions; or which treatment modality to offer, such as Cognitive Behavioral Therapy [CBT] or psychodynamic psychotherapy) and micro-level decisions (*which* specific treatment techniques and *how* they should be delivered). The *Structure* dimension characterizes the *method* of personalization, and the extent to which such a method is formalized or standardized. For simplicity, we locate these various approaches along a continuum from less structured to highly structured methods. At the less formal end of the continuum, *idiosyncratic* or *artisanal* approaches (Structure 1 or 3DP:S1) privilege a clinician's experience, intuition,

and judgment. Because the artisanal (Perlis, 2016) method of decision-making is highly variable from clinician to clinician, it is therefore less formalized and more difficult to replicate. Even if an expert clinician made consistently effective intuitive decisions that resulted in good treatment outcomes, such a method would not be easily accessible to other clinicians, unless it was somehow formalized and structured.

Along the *Structure* dimension, we also find *guidelines* (3DP:S2), which – in the field of psychotherapy – are usually developed by expert panels who survey and synthesize the current scientific evidence or consensus regarding best practices. According to the Institute of Medicine, clinical practice guidelines are systematically developed statements to help clinicians and patients make health care decisions under specific clinical circumstances (Field & Lohr, 1990). However, this assistance often takes a more narrative form, and many clinical practice guidelines are neither sufficiently structured nor intended to provide definitive answers or fixed protocols for clinicians to follow. More structure is offered by standardized *decision aids* (3DP:S3), which are defined as interventions that support clinicians and patients by making their decisions explicit, by providing information regarding options and associated benefits and harms, and by helping to align decisions with personal values (Stacey et al., 2017). As described in the development of the International Patient Decision Aids Standards checklist (Elwyn et al., 2006), decision aids can take the form of things like leaflets or videos that provide information about the available treatments, or they can be delivered as *algorithmic* tools such as interactive, rule- or tree-based decision aids. This illustrates that the labels along the Structure dimension are not all mutually exclusive categories, but instead function more as a spectrum of increasing specificity. Examples of this overlap are that many guidelines (3DP:S2) include decision-aids (3DP:S3), that a decision-rule (3DP:S4) is a special type of a decision-aid, and that some decision-rules are informed by statistical models (3DP:S5).

The final two categories along the Structure dimension both rely on algorithms to generate information for use in clinical decision-making. Algorithms lend standardization to the decision-making process, and their outputs (often a recommendation or prediction) are intended to be clear and precise. Thus, *algorithmic* approaches to personalization are highly formalized and structured, offering a greater degree of reproducibility. The defining feature of an algorithm is that it has clearly operationalized inputs and outputs, and a fixed, reproducible process for translating the inputs into the outputs.

One category of algorithms comprises *decision rules or decision trees* (3DP:S4), which can take the form of decision flow diagrams. When these are based on empirical data and research, they can be considered *data-informed*. However, it is important to highlight that the term algorithm, ubiquitous in the medical decision-making literature, should not be assumed to indicate that a model

is data-informed, data-driven, or derived from a statistical model. *Data-driven* models are those in which the relationships instantiated in the model attempt to represent, quantitatively, relationships that exist in the data from which they were derived (these distinctions are addressed in the subsequent section).

At the apex of the Structure dimension are algorithmic personalization approaches based on *statistical models* (3DP:S5). If well-developed, data-driven statistical models have the potential to improve the accuracy and effectiveness of clinical decision making. Although statistical models need not be derived from data (nor do they always aim to model real relationships), only data-driven statistical models will be discussed in this chapter.

The *structure* dimension is closely related to the sources of information that clinicians use to personalize treatment. These sources can include their clinical experience and intuition, available theories, empirical evidence, or quantitative data-driven models (see Figure 19.2 later in the chapter for more on sources of information).

The landscape of personalized psychotherapy approaches contains a wide variety of different combinations of features of the three dimensions of personalization. For example, there are: algorithmic statistical models designed to make decisions about which treatment package to offer a patient before the start of treatment (3DP:S5-L2-T1); decision aids based on theoretical models for choosing treatment strategies during treatment (3DP:S3-L3-T2); and algorithmic decision rules, designed to offer relapse prevention support in a targeted way, that identify patients at high risk of relapse after therapy (3DP:S4-L3-T3). This framework allows for the classification of a great diversity of approaches to personalization. In the next section, we will refer to the 3DP framework as we provide examples of early approaches to psychotherapy personalization.

Intuitive and Theoretical Approaches to Psychotherapy Personalization

Our historical overview of personalization approaches begins with a discussion of less formalized, *intuitive* approaches. This is followed by a discussion of *theoretical* approaches, which vary in their level of structure, and which typically have emerged from intuitive models as they become formalized and, in many cases, subjected to empirical tests.

Intuitive Approaches to Psychotherapy Personalization

Personalized psychotherapy has its roots in intuitive and theoretical models. *Idiosyncratic eclecticism* is an artisanal approach wherein therapists select treatment strategies and concepts based on their subjective appeal or their familiarity (Wheeler, 1993). Such an approach is underpinned by the belief that tacit, informal and intuitive clinical impressions are the most valid inputs available for guiding clinical decisions. Proponents of this approach in health care have argued that expert practitioners have an ability to know intuitively what will be the most effective way to intervene at any decision point in therapy (e.g., Benner & Tanner, 1987; Schon, 1984; Welsh & Lyons, 2001).

A common thread across these approaches is the notion that different patients, even if they share a diagnosis and other clinical features, require different therapeutic strategies. Also central to these approaches is the belief that, because patients' needs and goals change over time, therapists must tailor their approach to each patient, and to changes in a patient's needs and circumstances. Given that a therapist may draw upon an extensive array of techniques, a rationale is required to decide which techniques to apply for each case and at what specific point in time.

Despite its commonsense appeal, the wisdom of relying *solely* on clinical judgment is strongly questioned by a large body of accumulated evidence over the last 50 years. Authors of narrative reviews (Bell & Mellor, 2009; Dawes et al., 1989, 1993; Garb, 2005; Grove & Meehl, 1996; Meehl, 1954; Sawyer, 1966) and meta-analyses (Ægisdóttir et al., 2006; Grove et al., 2000) converge on the observation that clinical judgment and clinicians' prognostic assessments are highly variable and error-prone, are influenced by known biases (e.g., confirmation bias), and tend to be outperformed by formal/actuarial methods, such as those based on data-driven statistical models and decision rules. It therefore behooves clinicians to look for guidance from tools or systems beyond their own clinical judgment when making important treatment decisions (Magnavita & Lilienfeld, 2016).

Still, evidence-based tools are not widely accepted or used in routine psychotherapy practice (Gaudiano et al., 2011). Moreover, the influence of therapists' personality and epistemological beliefs, as well as their training and supervision experiences have been widely documented as determinants of their preferences for therapeutic orientations and selection of techniques (e.g., Arthur, 2001). Paradoxically, while the therapist may have the intention to personalize

therapy according to the patient's features, the process itself is likely influenced by a therapist's own features (e.g., personality, philosophical ideas, adherence to clinical intuition as a privileged source of information) and idiosyncratic circumstances (e.g., relationships with peers, the influence of particular supervisor). Therefore, idiosyncratic eclecticism fails to customize therapy for individuals based on formal theoretical models or data-driven models, instead privileging the therapist's attitudes, beliefs and preferences. This approach can also result in an unsystematic treatment process that is difficult to replicate and can impede therapists' ability to develop competence in any one approach or technique, which necessarily requires systematic repetition. Idiosyncratic eclecticism also makes it difficult to form an objective view about the effects of techniques, as it avoids the consistent application of an intervention and systematic observation of its effects across cases that would inform future decisions. Despite these shortcomings, eclecticism is frequently advocated in counseling and psychotherapy training courses (Wheeler, 1993) and is prominently endorsed in surveys of practicing therapists and psychologists (Norcross, 2005; Norcross & Prochaska, 1988).

The *pluralistic approach*, a close cousin of idiosyncratic eclecticism, acknowledges the idiosyncratic nature of clinicians' judgments and instead advocates that decisions should be guided less by therapists' judgments and more by patients' needs, wants and preferences (Cooper & McLeod, 2011). Therapists are thus encouraged to decide on the tasks of therapy based on collaborative discussions about patients' idiosyncratic goals and preferences about the style (e.g., more vs. less directive) and focus (e.g., past vs. present; building on strengths vs. addressing weaknesses) of therapy. In one version of this approach, the therapist relies on a structured inventory of preferences called the *Therapy Personalization Form* (TPF) to guide conversations about the initial selection of treatment strategy and style (Bowens & Cooper, 2012). Subsequent adaptations of therapy are accomplished by the completion and discussion of the TPF at regular review points. Compared to idiosyncratic eclecticism, the pluralistic approach is more systematic and mindful of research evidence (cf. Ong et al., 2020). However, it still relies mainly on clinical opinions and indirect evidence to guide its treatment selection method. For instance, the TPF approach was developed through a series of qualitative investigations of therapists' experiences of therapeutic dilemmas, and from the documentation of therapists' opinions about how to respond to patients' needs (Bowens & Cooper, 2012). To date, there is no evidence that personalization via TPF leads to superior treatment outcomes, relative to a standardized psychological treatment approach.

Theoretical Approaches to Psychotherapy Personalization

Toward the middle of the *Structure* dimension (Figure 19.1), we can find more systematic approaches that are guided by formal clinical assessment and theory. For example, Lazarus' multimodal therapy approach (2005) advocates the structured assessment of seven discrete and interactive modalities: behavior, affect, sensation, imagery, cognition, interpersonal relationships, and drugs/biology (BASIC ID), from which the therapist develops a BASIC ID profile. The profile, which informs a conceptualization of how these modalities interact to cause distress or dysfunction, serves as a formal blueprint to guide the selection of techniques to rectify problems in one or more domains. Furthermore, the sequence of techniques to be used is matched to the BASIC ID blueprint, which usually starts with one modality and moves sequentially to other modalities. Multimodal therapy is an example of *systematic technical eclecticism,* which is informed primarily by theory (in particular social-cognitive learning theory) but which also advocates techniques that are empirically supported. Notwithstanding its intuitive appeal, clinical parsimony, and the fact that it is premised on empirically supported theories, there is not a single experimental comparison of multimodal therapy versus inactive or active control conditions.

Another example of systematic technical eclecticism can be found in Beutler's *systematic treatment selection* (STS) approach (Beutler et al., 2005). In it, standardized questionnaires are used to assess four domains: functional impairment, coping style, reactance/resistance, and distress or emotional arousal. The resulting profile is used as a decision aid (3DP:S3) to prescribe specific techniques that are believed to be the most appropriate for a patient with that profile. An assessment of functional impairment is used to determine the intensity (e.g., frequency/duration of sessions, combination of group, individual, pharmacological interventions) of treatment. The assessment of coping style (internalizing vs. externalizing style) informs the selection of techniques (3DP:L3) that emphasize insight and activation of emotions (for internalizers) versus those that focus on practical skill-building and systematic feedback (for externalizers). Patients are also assessed according to reactance/resistance, which in turn guides the therapist to adopt a less directive approach for highly resistant patients and a more directive approach for those who are less resistant (3DP:L4). Finally, the patient's level of distress or emotional arousal is monitored to guide the selection of techniques that may serve to optimally modulate emotions. For example, strategies to promote cognitive dissonance may be selected in cases where blunted

affect may interfere with motivation and engagement, whereas for patients with highly aroused emotions strategies might aim to contain and soothe their distress.

Integrative Approaches

Integrative approaches are said to emerge in two ways: through the theoretical integration of two or more approaches or through the assimilation of other approaches into one's practice over time (Cooper & McLeod, 2011). Some examples of theoretical integration include cognitive analytic therapy (Ryle et al., 1990) and cognitive behavioral analysis system of psychotherapy (CBASP, McCullough Jr, 2003), both of which draw from cognitive and psychodynamic theories. Similarly, both dialectic behavioral therapy (DBT), which was developed by Linehan (1993) to treat suicidal patients, and mindfulness-based cognitive therapy (MBCT), which was developed by Segal, Williams and Teasdale (2001) as a relapse prevention strategy for individuals with major depressive disorder (MDD), demonstrate the gradual assimilation of meditation practices and emotion regulation strategies into a CBT framework. Integrative approaches do not always emerge with the explicit goal to personalize treatment for individuals, but the personalization imperative is often implicit in the rationale for theoretical or technical integration. The above cited therapies, for example, were born of the authors' dissatisfaction with then-current approaches and their apparent lack of efficacy/effectiveness with specific hard to treat groups: patients with chronic affective disorders, suicidality, and severe interpersonal difficulties. Thus, integrative treatments are often claimed to be well-suited for specific subgroups of the clinical population, such as those with dysthymia (CBASP), recurrent depression (MBCT), or borderline personality disorder and suicidality (DBT). In this way, integrative approaches often coalesce into standardized interventions, which are then selected for subgroups of patients who are assumed to be well-matched to the intervention, thus blending into the *theoretical* approach to personalization. Many of these formalized integrative treatments have been evaluated in RCTs and are now evidence-based interventions for specific patient populations (e.g., Piet & Hougaard, 2011).

Matching Treatments to Disorders

Indeed, one of the most prominent ways to approach the "what works for whom" question is to prescribe therapies that have empirical support for the treatment of specific clinical problems and populations. The vast majority of research regarding treatment selection over the past half-century has taken the form of identifying the best mapping of disorders to treatment packages, many of which were developed for specific disorders or patient populations. For example, prolonged exposure (PE) was developed by Foa to treat patients with post-

traumatic stress disorder (PTSD; Foa & Rothbaum, 1998). Largely founded on the controlled trials and meta-analytic evidence base, the disorder-matching approach has yielded different frameworks for treatment selection, where personalization mostly involves matching an evidence-based therapy to the patient's principal presenting problem or diagnosis. When algorithmic, this information can be used in a decision rule (e.g., if borderline personality disorder with suicidality, provide DBT; if PTSD, provide prolonged exposure; if obsessive-compulsive disorder [OCD], provide exposure and response prevention, etc.). If not a formal algorithm, as is often the case in guidelines, this information can take the form of a decision aid. Typically, the therapist and patient collaborate in reviewing initial diagnostic assessment results, they identify the principal problem(s), and this guides the selection of a disorder-specific treatment package or series of modules (i.e., discrete treatment tasks for the target problems).

This approach is endorsed by authoritative guidelines such as those published by the American Psychological Association's task force on the promotion and dissemination of psychological procedures (Chambless & Hollon, 1998; Chambless & Ollendick, 2001; Tolin et al., 2015), the UK National Institute for Health and Care Excellence (National Collaborating Centre for Mental Health [NCCMH], 2011), the Canadian Psychological Association (2012), and the Royal Australian and New Zealand College of Psychiatrists (Malhi et al., 2015). Such guidelines commonly advocate the prescription of first-line or empirically supported therapies, followed by "probably efficacious" treatments as second-line options when first-line treatments are ineffective, unavailable, or unacceptable to the individual. Criteria that must be met to be considered an empirically supported therapy usually include support by at least two well-conducted experimental studies, ideally carried out by different research groups, in clearly specified clinical samples, targeting specific outcome domains and supported by standardized treatment manuals (Chambless & Hollon, 1998; Tolin et al., 2015).

Guidelines cited above have produced diagnostic treatment matching systems for a range of common mental disorders such as depression, generalized anxiety disorder, obsessive-compulsive disorder, social anxiety disorder, specific phobias, panic disorder, substance use disorders, eating disorders, and other conditions. The integrative therapy approach described above has also tended to become assimilated into this algorithmic diagnosis-based treatment matching scheme. For example, MBCT has been recommended as a relapse prevention intervention, specifically for patients with recurrent major depression in remission who have experienced at least three previous episodes (Piet & Hougaard, 2011). In another example, DBT has been recommended to reduce the risk of self-harm in patients with borderline personality disorder, within a structured program including individual therapy, group-based skills training and access to support between sessions (NCCMH, 2009).

Notwithstanding its prominence in current practice guidelines, the disorder-matching approach faces a practical challenge from the proliferation of different treatment approaches, which has resulted in an abundance of evidence-based interventions for many of the common mental disorders. For example, prolonged exposure, cognitive-processing therapy (CPT), trauma-focused CBT, and eye movement desensitization and reprocessing (EMDR) are all evidence-based interventions for PTSD, and a variety of pharmacological and psychological interventions, including CBT, behavioral activation, interpersonal therapy (IPT), psychodynamic therapy, and emotion-focused therapy, are considered evidence-based treatments for depression. The disorder-matching treatment selection scheme has also been criticized on philosophical and methodological grounds. Some fundamental questions about the empirically supported therapy paradigm and approach to treatment selection concern the issues of generalizability, heterogeneity and diagnostic specificity. Given the rigorous selection criteria imposed in RCTs to safeguard internal validity, trial participants may not be representative of ordinary treatment-seeking samples in routine care. For example, people with certain comorbidities (e.g., substance use disorders, personality disorders, chronic illnesses) are often excluded, and the limited representation of people from minority racial and ethnic groups or disadvantaged socioeconomic backgrounds may limit the generalizability of the results of some RCTs to routine care settings (Henrich et al., 2010). This matter is accentuated by evidence that these very sources of heterogeneity (e.g., comorbidity) are known to be common. For example, Kessler et al. (2005) reported that only 50% of people with diagnosed mental disorders in an epidemiological survey met criteria for a single disorder, with the rest meeting criteria for two or more comorbid disorders. Furthermore, these sources of heterogeneity are relevant to patients' prognoses. For instance, systematic reviews indicate that comorbid personality disorders (Newton-Howes et al., 2014), comorbid chronic illnesses (Dickens et al., 2013), and socioeconomic deprivation (Finegan et al., 2018) are associated with poorer psychological treatment outcomes. This evidence partly supports Garfield's (1996) argument that overly focusing on diagnosis as a privileged criterion for treatment selection minimizes the influence of other sources of information that may be highly relevant to the outcome of a given treatment. Wampold and Imel (2015) argue that the many RCTs that generally find equivalent effects when comparing alternative forms of psychological treatment challenge the predominance of diagnostic classification as the sole criteria for treatment selection. While some have interpreted this latter observation as evidence that all *bona fide* therapies work through common factors, it is also plausible that different treatment techniques or packages work differentially for patients with specific attributes that may be unrelated to diagnosis (e.g., coping style and reactance, as described earlier). These differential response phenomena are known as aptitude-by-treatment interactions (ATI, Cronbach, 1957; Cronbach & Snow, 1977). Insofar as these phenomena are present, our ability to understand what works for whom will

be limited by recruiting homogeneous samples, focusing primarily on diagnosis, and examining clinical effects at an aggregated level.

Other Theory-Based Approaches to Treatment Selection and Adaptation

A common way in which psychotherapy researchers have sought to identify the treatments that are most effective for specific patients has been through the development of theory-driven treatment matching approaches. A citation classic in this literature is Project MATCH, a large (N = 1726) RCT designed to test *a priori* matching hypotheses about subgroups of patients with alcohol use problems who might respond differentially to three treatments: CBT, motivational enhancement therapy, and 12-step facilitation (Project MATCH Research Group, 1997). The results offered partial support for 4 out of 20 theoretically informed client characteristics (psychiatric severity, anger, support for drinking, and alcohol dependence). However, the significance of interaction effects varied across assessment time-points and the authors concluded that treatment matching did not substantially improve patient outcomes (Project MATCH Research Group, 1998).

In the area of depression treatment, Miller et al. (2005) randomly assigned N = 121 patients to receive either their matched or nonmatched treatment across four conditions: (1) pharmacotherapy; (2) combined pharmacotherapy and cognitive therapy; (3) combined pharmacotherapy and family therapy; and (4) a combination of all three treatments. The personalization algorithm took the form of a decision rule that used specific cutoffs on self-report questionnaires and an interviewer-rated scale collected pretreatment to indicate both the optimal level of care *and* treatment type (3DP:S4-L1+L2-T1). This matching system was constructed based on a deficit remediation theoretical model that hypothesized that patients with deficits in a specific area (cognitive distortions, family impairment) would benefit most from a treatment that directly addressed those deficits. The findings were mixed: adequately matched patients had better treatment outcomes compared to mismatched patients on some measures (Hamilton Rating Scale for Depression change, % improved) but not others (Beck Depression Inventory; suicidal ideation; symptom remission). Cheavens et al. (2012) provide another example of a similar approach to personalization based on the capitalization versus compensation model. Patients (N = 34) were assessed on therapy skill use across four domains (behavioral, cognitive, interpersonal, and mindfulness) using a semi-structured interview at baseline. A panel of evaluators and supervisors made consensus determinations about each patient's two relative strength domains and two relative deficit domains, and the patients were then randomized to receive the two modules that could capitalize on their existing strengths (capitalization), or to receive the two modules that could remediate their skill deficits (compensation). Their comparison of these two theory-based personalization approaches for

selecting technique-focused modules (3DP:S4-L3-T1) found evidence for early and sustained advantages in depression symptom reductions for those in the approach based on capitalization (Cheavens et al., 2012).

In another prospective trial, Watzke et al. (2010) randomized individuals (*N* = 291) initiating inpatient psychotherapy (either CBT or psychodynamic therapy) to either theoretically informed systematic treatment selection (distinct from Beutler's STS described earlier) or to random allocation. Treatment matching in the experimental group was carried out by multi-disciplinary team consensus, using a theoretical model whereby patients' goals, diagnoses, and other factors (preference, motivation, cognitive and interpersonal deficits, etc.) determined which treatment package they received (3DP:S3-L2-T1). Goals that were related to insight and emotion tended to favor a match with psychodynamic therapy, whereas goals that related to coping and behavior change favored a match with CBT. Patients with an anxiety disorder, eating disorder or post-traumatic stress disorder, as well those with a low level of cognitive reflection were also mostly matched with CBT. The main results found no general effect of systematic treatment selection, although secondary analyses suggested that, within those who received psychodynamic therapy, matched cases had better long-term outcomes than randomized cases.

Overall, *theoretical* models of treatment matching such as those described above have tended to yield mixed and inconclusive findings, which reflects the pattern observed for the vast number of aptitude-by-treatment interaction studies conducted over the last four decades (Dance & Neufeld, 1988; Smith & Sechrest, 1991; Snow, 1991). Such mixed and unreplicated findings could be related to methodological limitations such as small sample sizes and the predominant testing of single-moderator hypotheses. Recent studies have been moving toward data-driven multivariable approaches, where the combined influence of multiple attributes is examined in order to identify cases that respond differentially to alternative forms of treatment. These approaches will be covered in more detail in the subsequent sections of this chapter.

Another approach to personalization involves the adaptation of standard treatments with the goal of increasing their acceptability or effectiveness. Some have proposed that effective treatment adaptation can be achieved by expert psychotherapists using their clinical judgment and clinical case formulations (Betan & Binder, 2010; Eells et al., 2005; Persons, 1991; Smits et al., 2019). However, there is little experimental support for the supposed advantage of treatment adaptation guided by clinical intuition over standardized and protocol-driven interventions. Compared to intuitive models, theoretical models for treatment adaptation have gained some empirical support. According to psychological theory, therapy is more likely to be acceptable and effective if the rationale, concepts and delivery are reasonably well-aligned to a patient's worldview and culture (Cloitre et al., 2020). Significant meta-analytic support for

this claim exists (Benish et al., 2011; Chowdhary et al., 2014; Hall et al., 2016). One such meta-analysis by Soto and colleagues (2018) examined 99 studies (N = 13,813) and found that culturally-adapted interventions resulted in better treatment outcomes compared to standard interventions. In these studies, the most frequent adaptations involved providing treatment in the patient's preferred language, explicit mention of cultural content/values in treatment, and matching patients with therapists of a similar ethnic background. However, meta-analytic evidence regarding the impact of such adaptation is mixed (Captari et al., 2018; Cuijpers, Karyotaki, Reijnders, Purgato et al., 2018). For example, a meta-analysis by Cabral and Smith (2011) found almost no benefit to matching clients with therapists based on race or ethnicity.

Another theory in psychotherapy is that collaboration and the therapeutic alliance can be enhanced by accommodating a patient's preferences, thus resulting in better treatment adherence and therapeutic outcomes. A meta-analysis of 29 randomized clinical trials (N = 5294) found that providing patients with their preferred psychosocial mental health treatment was associated with lower dropout and greater alliance scores, but not associated with depression and anxiety outcomes, remission, global outcomes, attendance, or treatment satisfaction (Windle et al., 2019). Patients' expectations and preferences for a particular treatment approach could, in turn, be influenced by their interaction with the therapist, or through preparatory interventions such as role inductions and motivational interviewing (Ogrodniczuk et al., 2005; Walitzer et al., 1999). The conceptual and statistical challenges that can arise when investigating patient preference in the context of randomized clinical trials require careful consideration (Gemmell & Dunn, 2011; Walter et al., 2017). Although preference accommodation at the level of selecting interventions seems to be supported by empirical research (Zoellner et al., 2019), the benefits of accommodating preferences about therapy style and techniques require further empirical investigation through comparisons against standardized treatment models with or without adequate preparatory strategies.

Summary: The Road to Precision Starts with Trial and Error

In this section, we have endeavored to present a historical overview of intuitive and theoretical approaches developed with the aim to personalize and ultimately improve the effectiveness of psychotherapy. Rather than an exhaustive review of studies, we have chosen to provide a developmental perspective on the evolution of concepts, methods and trends in the field, supported by a selection of illustrative examples. The personalization imperative has often been approached from an intuitive perspective, based on the notion that clinical expertise is

sufficient to meet the needs of patients encountered in practice. Yet, evidence collected over 50 years convincingly demonstrates that clinical judgment is highly variable and error-prone. Nevertheless, clinical observation and speculation have paved the way for the development of theories that generate formal hypotheses about what works for whom and thus are more amenable to validation through experimental tests. In recent decades, the advent of empirically supported treatments and the growth of the clinical trials literature has led to the development of theoretical approaches for treatment selection. Today, the most common forms of treatment selection applied in routine care clinical practice rely on guidelines based on the empirically supported therapy literature and disorder-matching. Other theoretical approaches that aim to reconcile personalization with current clinical practice guidelines have focused on ways to adapt empirically supported therapies to the features of individual patients. Of these approaches, cultural adaptation has accrued the most empirical support. Overall, intuition and theory have paved the way for the development of more structured and precise *algorithmic* approaches, which are the focus of the remaining sections of this chapter. This historical development, moving from intuitive toward theoretical, data-informed and data-driven approaches, is represented in Figure 19.2.

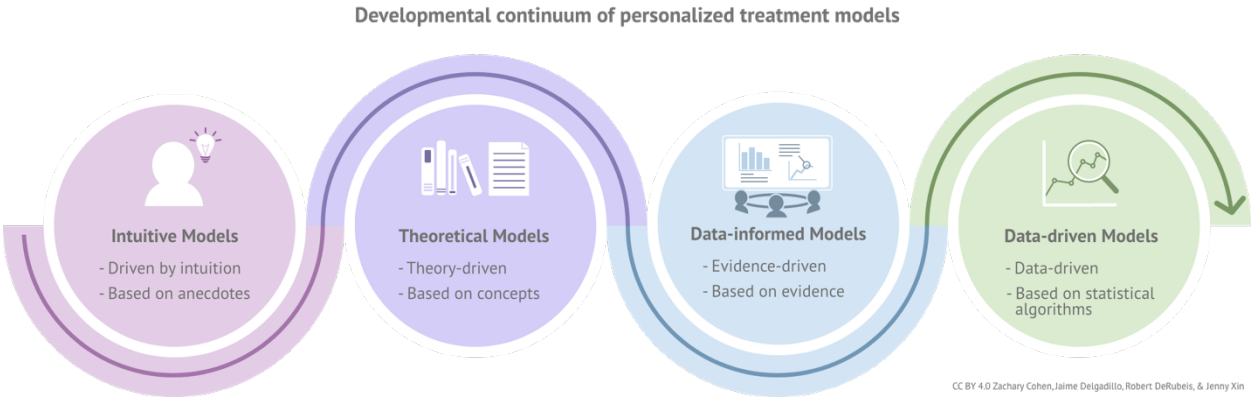

Figure 19.2. Sources of information along a developmental continuum of personalized treatment approaches.

Contemporary Applications of Personalization in Psychotherapy

Data-driven approaches to personalization are a recent development in the field of psychotherapy. The basic premise of data-driven treatment personalization is relatively simple and theoretically sound: if you can predict something relevant about a treatment decision that contains uncertainty, this information could be

used to increase the likelihood of making better decisions, which should improve outcomes in the aggregate. These approaches, which are the foundation for the highly structured approaches from the upper end of the *Structure* continuum depicted in Figure 19.1, have several distinctive features: they are developed using observed data (e.g., outcomes from psychotherapy patients); they are developed using statistical analyses; and their outputs (e.g., treatment recommendations or prognostic information) are generated by an algorithm. As such, the decisions that are guided by data-driven approaches are highly standardized, and therefore reproducible, in contrast to decisions derived from intuitive approaches. In this section, we will describe several data-driven approaches, organized according to the *Level of Intervention* domain outlined in Figure 19.1, which aim to inform decisions about how to personalize care by selecting the right treatment intensity, package, or component, or by adjusting therapy delivery or interaction style.

Selecting the Intensity of Psychotherapeutic Interventions in Stratified Care

Across medicine, it is common for interventions to vary in their level of intensity and cost. For example, distinctions are made between primary care (routine interventions), secondary care (specialist interventions), and tertiary or hospital-based care (more invasive or costly interventions). Similarly, a specific treatment modality can vary in intensity. For example, in pharmacotherapy, a single medicine can be prescribed or it can be augmented or combined with other medicines. In psychotherapy as well, intensity can vary from lower to higher intensity interventions. Low intensity interventions are often provided for less severe or non-chronic cases of common mental disorders (NCCMH, 2011). These interventions include brief (i.e., eight or fewer sessions), low-cost, psychoeducational interventions such a computerized CBT, bibliotherapy, and group-based treatments. Such interventions have garnered considerable empirical support in the treatment of depression and anxiety disorders (Coull & Morris, 2011). In contrast, high intensity interventions include empirically supported individual psychotherapies such as those described earlier in this chapter (e.g., CBT, IPT, EMDR, PE, CPT etc.), as well as combination treatments – psychotherapy plus pharmacotherapy – which can be twice as lengthy and costly as low intensity treatment options. Given some evidence that low and high intensity psychological interventions can be equally effective *on average* (e.g., internet-based vs. face-to-face CBT; Carlbring et al., 2018), stepped care has been recommended in clinical guidelines as a cost-effective way to provide low and high intensity treatments in a sequential way (NCCMH, 2011). Stepped care has a well-established evidence base (Firth et al., 2015), and it is a standard strategy for mental health care delivery in countries like the United Kingdom (Clark, 2018).

Recent work has highlighted how the identification of different patient subgroups could inform comparisons of therapies of different strengths. Building on work in which they had proposed the *patient response profiles* framework (DeRubeis, Gelfand et al., 2014), Gelfand and colleagues presented simulations that demonstrate how, depending on the mix of patient types, a comparison between a stronger and a weaker treatment can either reflect the actual difference in strength or can make it appear that there is little if any difference between the two treatments (Gelfand et al., under review). Their analysis highlights the importance of identifying, in advance if possible, those who are as likely to benefit from a less intensive treatment as from a stronger treatment, and those for whom a greater benefit is expected in the stronger treatment. Recent work has extended these concepts to comparisons of diagnosis-specific and transdiagnostic interventions (Dalgleish et al., 2020).

In recent years, researchers have developed data-driven models to identify different phenotypes or subgroups of patients with specific characteristics who may vary in their need for higher intensity treatments for their depression or anxiety. In a large (N = 16,636) naturalistic cohort study of patients who accessed stepped care interventions, Saunders et al. (2016) applied latent cluster analysis to identify eight subgroups of patients who shared similar pretreatment clinical and demographic features. The authors found that patients classified into three of the profiles tended to have a higher probability of depression symptom recovery in high intensity relative to low intensity treatments, whereas those in the other five profiles did not. Using data from an independent sample of 44,905 patients, Saunders et al. (2020) replicated many of these findings using a latent variable mixture modeling approach that leveraged the eight profiles previously identified in Saunders et al. (2016). In addition to observing that a subset of the profiles were associated with differential outcomes between low and high intensity therapies, they examined differences in outcome between different types of high intensity therapies (CBT and Counseling) and identified profiles that were associated with differential likelihood of reliable recovery in one specific intervention versus the other.

Delgadillo and colleagues (2016), applying bootstrapped logistic regression methods to data obtained from a stepped care psychological treatment service, identified robust predictors of depression treatment outcomes. Based on the regression weights associated with these predictors, they developed the Leeds Risk Index (LRI), which classifies patients as having low, moderate or high predicted risk of enduring depression symptoms after treatment. In a prospective field-test of the index, therapists used the LRI to guide the assignment of patients (N = 157) to either low or high intensity treatments (Delgadillo, Appleby et al., 2020). The pattern of outcomes in this uncontrolled open trial of LRI-informed allocation was favorable in comparison to a case-control matched sample drawn from archival clinical records that had undergone stepped care.

Recent studies have applied penalized regression approaches to develop prognostic indices, which combine pretreatment patient features in a way that can help rank cases according to their expected outcomes (e.g., a continuum ranging from excellent to poor predicted treatment response). Using data ($N = 622$) from a three-arm randomized controlled trial (treatment as usual [TAU] vs. brief therapy vs. CBT), Lorenzo-Luaces et al. (2017) combined several pretreatment patient characteristics using a regularized regression equation (bootstrapped LASSO technique) to develop a depression prognostic index. Using this index, they observed that the 75% of patients with the best prognoses responded similarly to all treatments. However, among the 25% of cases with poorer prognoses, recovery rates were substantially higher in the CBT condition (60%) than in TAU (39%) or brief therapy (44%). Using a similar modeling approach in a naturalistic cohort study ($N = 1512$) of stepped care treatments, Delgadillo et al. (2017) developed a depression prognostic index which was used to classify cases into two groups: complex cases (28% of cases with the poorest expected prognosis) and standard cases (the majority). Using a split-halves cross-validation procedure, they observed that complex cases in the test sample had higher depression recovery rates when they were assigned directly to high intensity therapy rather than starting with low intensity psychological interventions. This advantage was large and statistically significant for the complex cases, whereas it was small and nonsignificant for the standard cases.

The algorithm developed by Delgadillo et al. (2017) has been implemented in a large ($N = 951$) multiservice cluster randomized trial (Delgadillo et al., under review), making it the first prospective test of the utility of a data-driven treatment recommendation algorithm for stratification in psychotherapy. Depression recovery rates were higher in patients whose assignment was informed by the algorithm's recommendation, relative to the patients who were allocated according to usual stepped care. The findings from this study are encouraging regarding the potential for the development, implementation, and dissemination of data-driven treatment selection approaches.

Selecting Treatment Packages

Systematic reviews of evidence-based interventions within specific patient populations have generally supported the conclusion that the *average* treatment effects of most evidence-based interventions are not different from one another. For example, Cuijpers and colleagues (2020) conducted a network meta-analysis to examine 101 studies of treatments for depression, comprising 11,910 patients, and found no difference in rates of response or remission between patients who received antidepressants or psychotherapy. However, they also found significant heterogeneity of observed treatment responses, reflecting the fact that some patients show significant improvement whereas others evidence little-to-no

improvement, or even deterioration. Thus, one of the most basic and essential questions that precision medicine approaches can answer in mental health is *which* of the available evidence-based treatment packages is likely to result in the best outcome for a given individual (3DP:L2). This is the core question for personalized treatment selection (Cohen & DeRubeis, 2018).

Several data-driven approaches to treatment selection have been described in the literature. What most of these approaches have in common is the goal of generating an index of the expected differential magnitude – or likelihood – of response between two (or more) available treatment packages. All of these approaches use statistical modeling to capture the collective predictive signal from multiple pretreatment variables. If this signal is strong enough, the index derived from it can be used to inform, for each new patient, a treatment recommendation. It is understood that, for many patients, even a powerful model will predict little if any differential benefit from one treatment versus another. But a powerful model will be able to inform treatment decisions in such a way that substantially better outcomes are achieved, in the aggregate, relative to random assignment (or one of the other comparisons described below in the next section)(Kapelner et al., 2021).

One of these approaches, proposed by DeRubeis, Cohen and colleagues (2014), yields for any individual a *Personalized Advantage Index* (PAI), which is a signed score that indicates which of two interventions is expected to result in a better outcome. The magnitude of the index reflects the strength of the expectation. DeRubeis, Cohen et al. (2014) first demonstrated the PAI approach using data from a randomized comparison of CBT versus pharmacotherapy (paroxetine) in outpatients with major depressive disorder. Numerous studies and research groups have applied variations of this PAI approach to examine if subgroups of patients can be identified who respond differentially to: CBT vs. interpersonal psychotherapy (Huibers et al., 2015; van Bronswijk, Bruijniks et al., 2020; van Bronswijk et al., 2019); trauma-focused CBT vs. EMDR (Deisenhofer et al., 2018); prolonged exposure vs. cognitive processing therapy (Keefe et al., 2018); group CBT vs. transdiagnostic group CBT (Eskildsen et al., 2020); group CBT vs. a group integrative positive psychological intervention (Lopez-Gomez et al., 2019); CBT vs. CBT with exposure and emotion-focused elements (Friedl, Berger et al., 2020); usual face-to-face CBT vs. blended face-to-face plus internet CBT (Friedl, Krieger et al., 2020); CBT vs. psychodynamic therapy (Cohen et al., 2020; Schwartz et al., 2020); CBT vs. counselling for depression (Delgadillo & Gonzalez Salas Duhne, 2020); ADM (sertraline) vs. placebo (Webb, Trivedi et al., 2019); continuation-phase cognitive therapy vs. antidepressant medication (fluoxetine, Vittengl et al., 2017); and DBT vs. general psychiatric management (Keefe et al., 2020).

Other approaches have been used to develop data-driven treatment selection models. Nearest neighbors modeling (Lutz et al., 2006) identifies, for each treatment, a subset of patients who are most similar to a given index patient on a

set of variables (i.e., the index patient's "nearest neighbors"). The recommendation for the index patient is derived by calculating the difference in average outcome scores observed in one treatment versus the other, among the index patient's nearest neighbors. Other methods with similar goals are described by Kraemer (2013) – the M* approach, implemented by Wallace et al. (2013) and Petkova et al.'s (2017) composite moderator approach, as demonstrated by Cloitre and colleagues (2016) as well as in a recent tutorial (Petkova et al., 2019).

An index of predicted differential benefit can also be derived from two (or more) separate prognostic models, each of which is constructed using only data from patients who were treated with one of the interventions (e.g., Deisenhofer et al., 2018; Delgadillo & Gonzalez Salas Duhne, 2020; Furukawa, Debray et al., 2020; Kessler et al., 2017). The difference between the patient's predicted outcome in each treatment forms the basis for the treatment recommendation.

Other approaches for predicting differential benefit exist. Methodological development in statistical prediction algorithms is ongoing, and the application of novel approaches such as the super learner ensemble machine learning method (van der Laan et al., 2007) promises to provide increasingly powerful predictive information to support treatment personalization. When more than two options are under consideration, the statistical and practical complexity of treatment selection increases. For examples of such developments, see Zilcha-Mano et al. (2016) and Furukawa, Debray et al. (2020).

Selecting Treatment Components or Techniques

Another approach to psychotherapy personalization involves the selection of treatment components or techniques in a way that would maximize improvement for patients with specific features (3DP:L3). Developing such a system, and defining an explicit decision-making process to select the type and sequence of treatment components, has been the aim of recent investigations. The advent of modular psychotherapies, such as the unified protocol (Barlow et al., 2017) and MATCH (Chorpita & Daleiden, 2009), presents a novel opportunity to formalize this approach, wherein a therapist constructs a modular treatment package that is meant to address a patient's principal symptoms and concerns. In the context of psychotherapy, Chorpita et al. (2005) defined modularity as breaking up complex activities into self-contained, smaller units that can function independently while also working together as part of a package.

For psychological treatments that can be modularized, two goals of personalized mental health research are to attempt to predict *which* components are going to be most important for an individual to receive, and to predict the optimal *order* in which components should be delivered. Importantly, this approach to personalization does not require that the treatment components

themselves be altered. Individually constructed modular treatments could maximize efficiency by ensuring that patients are provided with only the specific components they need, and could increase adherence by minimizing the likelihood of patients receiving irrelevant or contraindicated interventions. These aims are relevant for both face-to-face psychotherapy as well as technology-based interventions. They are especially relevant in the context of transdiagnostic treatments, and when creating treatment plans for patients with significant comorbidity who present with a variety of potential targets. Another attractive feature of this internal personalization approach is that, unlike selecting between therapeutic modalities that would likely be delivered by different providers, this approach allows the therapist to personalize treatment themselves, as opposed to (for example) having to refer their client to another provider.

Theory-based approaches for personalizing treatment by determining which treatment components to deliver have often failed to outperform non-personalized interventions. For example, Moritz et al. (2016) assessed patients with OCD (N = 89) on 14 different cognitive biases and dysfunctional coping styles and then randomized the patients to receive either the full version of a self-help metacognitive bibliotherapy or an individually adapted version that only included the specific chapters theoretically indicated to address their endorsed target areas. The authors found that this algorithmic approach to personalization (3DP:S4-L3-T1) did not outperform simple provision of the entire manual (Moritz et al., 2016). Another example of a theory-based approach is "The Toolbox," a web-based self-guided app recommendation service in which users could select one or more "well-being categories" and then one or more goals, and then would be provided with a selection of what were determined to be relevant apps and tools to browse and download (3DP:S4-L3-T1). The personalized recommendation algorithm was developed through a participatory design approach with young people and mental health professionals based on a theory of change and theories of reasoned and planned behavior (Antezana et al., 2015). Bidargaddi and colleagues (2017) evaluated this non–data-driven approach to personalization in a randomized trial that compared it against a four-week waitlist control and found no benefit on the primary outcome.

In contrast to many theory-based approaches, randomized comparisons of data-informed personalized modular psychotherapy against usual-care conditions have often reported favorable results (e.g., Chorpita et al., 2017; Weisz et al., 2012). In a sample of youths, Weisz and colleagues (2012) evaluated a data-informed personalized modular treatment (MATCH) that comprised procedures from each of three separate standard interventions for depression, anxiety, and conduct disorder. The RCT (N = 174) compared MATCH against standardized treatment (one of three manualized interventions) and a usual-care control arm, and found that the modular condition was associated with superior outcomes relative to standardized treatment and usual-care.

Mohr et al. (2019) provide an example of a data-driven approach to recommending treatment components (3DP:L3). The algorithm was developed in a large training sample ($N > 100{,}000$) primarily by modeling associations between client characteristics and differential engagement with the modules. The authors noted that the decision not to use differential change in clients' symptoms was largely driven by data availability (Personal Communication, Mohr & Schueller, 11/13/2019). However, given the high rate of dropout often observed in digital mental health (Cuijpers et al., 2019), maximizing engagement is a sensible goal. In their factorial RCT ($N = 301$), Mohr et al. (2019) reported that, relative to those who received no recommendations, those who received the weekly personalized recommendations evidenced greater engagement and larger improvements in depression, but not anxiety, symptoms. Unfortunately, given the lack of a control comparison (e.g., sham recommendations), the possibility remains that, instead of being due to the personalized nature of the recommendations, the improved outcomes were simply the result of receiving reminders to download and use the apps, and that non-personalized recommendations might have worked just as well (or better).

The personalization discussed so far has relied on *nomothetic* statistical approaches, which trade on similarities and differences between individuals. In nomothetic personalization, predictions of response to specific treatment strategies are made for individuals using data from a reference group of cases treated in the past and who have features that are similar to the target patients. *Idiographic* approaches (Fisher & Bosley, 2020; Fisher et al., 2018; Wright & Woods, 2020), on the other hand, focus on within-person patterns, rather than patterns that are observed across individuals. The construction of these person-specific statistical models requires intensive longitudinal measurement. A model developed from an individual's own data can then be used to inform decisions about specific treatment strategies for that individual (Fernandez et al., 2017; Rubel et al., 2018).

For example, Fisher et al. (2019) developed personalized treatment plans for 40 CBT patients who had completed daily symptom measures over a one-month pre-treatment period, using ecological momentary assessments. Patients' symptom profiles were examined using person-specific factor-analytic (P-technique) and dynamic factor modeling techniques, and this information guided the construction of individualized treatment plans which drew upon modules from the Unified Protocol for transdiagnostic treatment of emotional disorders (Barlow et al., 2011). The authors reported large pre-post treatment effect sizes ($g = 1.86$ in the intention-to-treat analysis) for this individualized treatment approach, although – as the authors note – these findings are from a small ($N = 32$ completers) uncontrolled open trial.

Nuances in the methodology of this interesting trial can help illustrate some of the complexity in the personalization literature that motivated our creation of the frameworks described in Figures 19.1–19.3. The trial (Fisher et al., 2019) was

split into two phases that demonstrate different ways in which data-driven approaches can inform treatment selection. For one phase, an expert panel of clinicians met to discuss each patient's ideographic modeling results and develop the personalized treatment plans (Figure 19.3, process #5). For another phase, a dynamic assessment and treatment algorithm (Fernandez et al., 2017) automatically translated each patient's ideographic modeling results into a personalized treatment plan (see Figure 19.3, process #6) using an item-module matching matrix. Of interest is that this matrix was constructed based on a predominantly theory-driven match between the different symptoms and CBT strategies in the modules believed to target such symptoms, instead of, for example, on empirical evidence of whether each module of the Unified Protocol (or the technique it contained) had been shown to impact one or more of the target symptoms (i.e., data-informed), or an actual quantification of the magnitude of the effect of each module on the targets (i.e., data-driven). Thus, the second personalization approach described by Fisher et al. (2019) is a statistical model (3DP:S5-L3-T1) that illustrates a blend of a data-driven approach (the ideographic modeling component) with a theoretical approach (the item-module matching matrix). These idiographic methods and related process-based therapies (e.g., Hofmann & Hayes, 2019) are promising new approaches for advancing data-driven psychotherapy personalization.

Adjusting Psychotherapy as It Unfolds

Another important way in which personalization approaches differ concerns *when* the personalization scheme is applied (the Time dimension of Figure 19.1). Thus far, we have focused on the decisions that are made prior to the initiation of treatment (3DP:T1). When decisions are made, instead, dynamically over the course of treatment, one can refer to this as *adjusting treatment* (3DP:T2). Much of this growing literature focuses on adaptive treatment strategies such as just-in-time adaptive interventions (JITAI), which have the goal of delivering specific components of an intervention package at the optimal time for each individual (Nahum-Shani et al., 2015, 2018). These approaches have great potential for personalizing treatment, especially in the context of digital interventions (Goldstein et al., 2017) and digital phenotyping of mobile devices (Bae et al., 2018).

One foundational practice that can support treatment adjustment is routine outcome monitoring, whereby a client's symptoms are tracked over time using validated psychometric measures (Lambert et al., 2018). Routine outcome monitoring is used in *feedback-informed treatment*, in which the clinician is notified as to whether their patient is "on track" or "not on track," a categorization that relies on a comparison of a patient's progress against expected progress, based on available normative data from clinical samples. Numerous systematic reviews

and meta-analyses demonstrate that dynamically adjusting the treatment plan and approach using feedback-informed treatment leads to improved psychotherapy outcomes. A detailed review of the empirical support for feedback-informed treatment is found in Chapter 4 of this *Handbook*.

Another way in which time-varying prediction approaches can be used to improve patient outcomes is by identifying when a patient is at a high risk for dropout, so that interventions to increase the likelihood of retention can be initiated (Lutz et al., 2019, 2020). For example, Forsell et al. (2019) constructed predictive models using only longitudinal symptom data from a large naturalistic sample ($N > 5000$) of patients in a 12-week program of therapist guided internet-delivered CBT for depression, panic, or social anxiety. They found that sufficiently accurate predictions (which they defined as a balanced accuracy of 65%) could be generated starting halfway through treatment. Lorimer and colleagues (2020) demonstrated how risk of relapse could be dynamically predicted across the 12-month period following successful acute treatment using a multivariable ensemble machine learning approach.

Levin and colleagues (2018) evaluated a data-informed (but not data-driven) personalized psychotherapy in a small RCT ($N = 69$) that compared four weeks of digital acceptance and commitment therapy (ACT) with skill coaching that was tailored based on ecological momentary assessment (EMA) against random skill coaching or no skill coaching. All participants completed EMA, and participants in both skills conditions were blinded to allocation. Those in the tailored condition received the ACT skill component linked to the highest-rated problem (where the linking was based on both theory and observed associations, derived from an earlier pilot study, between specific skill coaching sessions and specific outcome variables). Although the authors acknowledged that the study was underpowered to detect differences between the active conditions, they found that the personalized just-in-time adaptive intervention (3DP:S5-L4-T2) led to significant improvements in a variety of outcomes, relative to the random coaching and control conditions.

Smyth and Heron (2016) provide a clear example of the potential of data-informed decision-rules for personalizing treatment that occurs dynamically across treatment (3DP:S4-L4-T2). Participants ($N = 90$) were randomized to one of three conditions in which they completed EMA about stress and mood 3X-per-day: EMA-only, EMA plus tailored recommendations about micro-interventions for responding to stress and negative mood, or EMA plus semi-random reminders about using randomly selected micro-interventions. In the tailored condition, recommendations were provided when a participant's EMA indicated moderately-high or high stress or negative affect (based on >= 1 SD above their person-centered mean, which was determined based on a two-day lead-in where they only completed EMA). The specific intervention recommendations were based on participants' symptom reports in the EMA. Relative to the comparison conditions,

the data-informed JITAI resulted in lower self-reported negative affect, fewer reported stressful events, lower cortisol levels, better sleep quality, and lower levels of potentially problematic health behaviors (e.g., alcohol consumption).

Methodological and Statistical Issues for Personalized Psychotherapy

> *"All models are wrong, but some are useful."*
>
> – Famous aphorism on statistical modeling coined by George Box (Box & Draper, 1987).

The decision to implement a new approach for personalizing treatment that deviates from current evidence-based practices should be informed by evidence of its utility. A savvy consumer of the constantly evolving field of personalized mental health care will want to know how to evaluate the literature critically. In general, there are several stages in the development and validation of data-driven personalization models (Delgadillo & Lutz, 2020). Early stages involve the construction and testing of a clinical prediction model that is intended to inform decision-making. More advanced stages involve case studies of implementation, prospective field-tests (e.g., Delgadillo, Appleby et al., 2020), and clinical trials of the resulting decision support systems in clinical practice (e.g., Lutz et al., 2021). A personalization system must first be constructed in order for it to be evaluated, and it must be described in sufficient detail if it is to be implemented by those who did not develop it. In the following, we address these three features of personalization systems. We begin with considerations relevant to the evaluation of such systems, as understanding how an approach has been subjected to empirical tests and what is known about the value it can provide is, we believe, of utmost importance to the general consumer of research on personalization. Although it can be helpful to know and understand how an approach was developed, it is not required for an independent evaluation of its utility.

The following high-level overview of model construction and evaluation builds upon decades of work by groups across medicine and statistics. More extensive discussion and guidelines for prediction model development and validation are provided by the "Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis" (TRIPOD) system (Collins et al., 2015; Moons et al., 2015). The PROGRESS (PROGnosis RESearch Strategy) group has provided recommendations for predictive modeling in health care research (Hemingway et al., 2013; Hingorani et al., 2013; Riley et al., 2013; Steyerberg et al., 2013). The Predictive Approaches to Treatment effect

Heterogeneity (PATH) Statement gives specific guidance for using predictive modeling of heterogeneity of treatment effects to inform clinical practice (Kent, Paulus et al., 2020; Kent, Van Klaveren et al., 2020). Wherever possible, we will refer the reader to these and other resources for more comprehensive discussions of relevant issues (e.g., Riley, van der Windt et al., 2019; Steyerberg, 2019).

Evaluation

In this section, we review the variety of approaches that can be used to evaluate psychotherapy personalization systems and the data-driven prediction models on which they are sometimes based. We will discuss contemporary methods for the evaluation of approaches used to select treatment packages as well as those used to tailor treatment.

Evaluating the Performance of Data-Driven Models

A statistical model developed in a given dataset (often called a *training sample*) should be evaluated using held-out data to estimate key performance measures, such as accuracy, for the predictions (Jacobucci et al., 2020). Held-out data can also be used to gain insight into the potential utility of the recommendations that can be derived from a predictive model. When an external data-source is unavailable for use as a held-out test-sample, internal validation, via techniques such as cross-validation or bootstrapping, should be performed to generate an estimate of the model's ability to provide accurate and precise predictions in a new sample (Steyerberg et al., 2001). The split halves approach should be avoided (Steyerberg & Harrell, 2016). Cross-validation methods can also be used internally during model construction, such as when selecting variables or tuning model hyperparameters. When estimates of a model's performance are derived from internal validation, failure to incorporate in the cross-validation all the steps needed to generate the model will lead to biased (overoptimistic) estimates of model performance (Steyerberg et al., 2003). The steps include variable selection, imputation of missing data, and weight setting. Even when all pertinent procedures have been included during cross-validation, the estimate of model accuracy can still be biased if the training sample is small (Varoquaux, 2018).

A variety of factors must be considered when estimating the sample sizes needed to develop (Riley, Snell et al., 2019a, 2019b; van Smeden et al., 2019) and validate (Archer et al., 2020) prediction models. Riley and colleagues (2020) provide comprehensive guidance on sample size issues, including an R package that implements their recommended procedure.

The generalizability of a model's predictions will likely be limited if it is applied in new cases or services that differ considerably to those used to train the model. It cannot be known with certainty, in advance, what characteristics of patients, treatments, or providers, will affect generalizability, so the best we can do is to apply our knowledge of the domains and adjust our expectations of the utility of the model in the test sample accordingly. For example, it would be unsurprising if a model developed in a study of university counseling patients generalized poorly to an inpatient care population. Prospective randomized experiments that use the same assessment and treatment procedures as were used to develop the model avoid many of the limitations encountered in tests that use preexisting data or data from contexts that differ from the model's development. We refer to this as *context specificity*: a prediction model is only expected to generalize to new samples that are sufficiently similar (e.g., in their features and treatment context).

An important aspect of the development and evaluation of statistical models for treatment personalization is *calibration*, which refers to the extent to which a model's predictions align with the true observed values (Alba et al., 2017; Steyerberg et al., 2010). Van Calster and colleagues (2019) highlight the potential dangers of – as well as advice on how to avoid – poorly calibrated predictive models in clinical decision-making. Huang and colleagues (2020) provide an extensive tutorial aimed at clinical researchers, complete with R code demonstrating calibration of predictive models for treatment personalization.

Another way in which prediction performance can be evaluated is through prospective designs. For example, Symons and colleagues (2020) developed a set of multivariable prediction models using data from patients ($N = 1016$) who had received a CBT-based intervention for alcohol dependence. They then generated prospective predictions for 220 new patients using these models and compared them against predictions made by 10 clinical psychologists, and observed superior accuracy and calibration for the model-based predictions.

In summary, numerous methodological choices can affect the performance and clinical utility of prediction models. Ideally, clinical prediction models are constructed using sample sizes large enough to provide adequate power, missing data are addressed properly, internal and external cross-validation methods are used appropriately, and evaluations of generalizability are conducted on independent test samples.

Evaluating the Potential Clinical Utility of Treatment Personalization Approaches

When a treatment personalization approach relies on a prediction model, accurate and precise predictions can reasonably be assumed to be a necessary condition for the success of the system, but they should not be assumed to be

sufficient. Thus, although it is important to use data and procedures that maximize the chance of obtaining a valid estimate of a prediction model's performance (e.g., accuracy), the question that will reveal the real-world utility of a model in a given context is: "Can the application of this treatment personalization approach, informed by the model's predictions, improve patients' outcomes, relative to the way in which the relevant decisions are currently made?" Answering that question requires the consideration of a multitude of factors beyond predictive accuracy.

Figure 19.3 depicts a variety of ways in which decision-making processes can translate information about the patient into specific treatment-relevant decisions. The first column of *The Information-to-Decision Translational Framework* aligns with the *Structure* dimension of the 3DP framework (see Figure 19.1). The second and third columns describe the outputs, which can be quantitative (e.g., predicted post-treatment symptom score) or qualitative (e.g., "psychotherapy is the optimal treatment for this client"). The fourth column (human filter) illustrates that information from a treatment personalization approach either leads directly to changes in the intervention for the client (i.e., without review by a human, as in #4 and #6), or is used by the clinician, client, or both in the form of a shared decision-making process to determine how to personalize treatment. Tests of treatment personalization systems that employ a human filter must be able to account for the fact that patients or clinicians can override the recommendation, thus affecting the extent to which the evaluation still reflects the quality of the quantitative or qualitative information that was generated.

Kraepelien and colleagues' (2019) study of a personalization approach highlights this issue. The authors describe how, after generating the personalized treatment plan, "The therapist would then present the patient with two recommended problem areas to work with, but the patient had the final word and could change to the two preferred problem areas of his or her choice. There was no measure of how often the patient chose to focus on other problem areas than the recommended" (Kraepelien et al., 2019, p. 3). As the patients were allowed to disregard the recommendations, and because there was no record of how frequently this occurred, a post-hoc analysis exploring the outcomes for the subset of patients who decided to follow the recommendations was not possible.

Figure 19.3 was created to be illustrative, and thus does not provide a comprehensive mapping of all possible scenarios. For example, Figure 19.3's process #5 does not capture perfectly the first approach used by Fisher and colleagues (2019) that was described earlier in this chapter, in which quantitative outputs generated by statistical models were sent directly to an expert panel of clinicians who then developed the personalized treatment plan (without the quantitative outputs first being translated into an explicit recommendation).

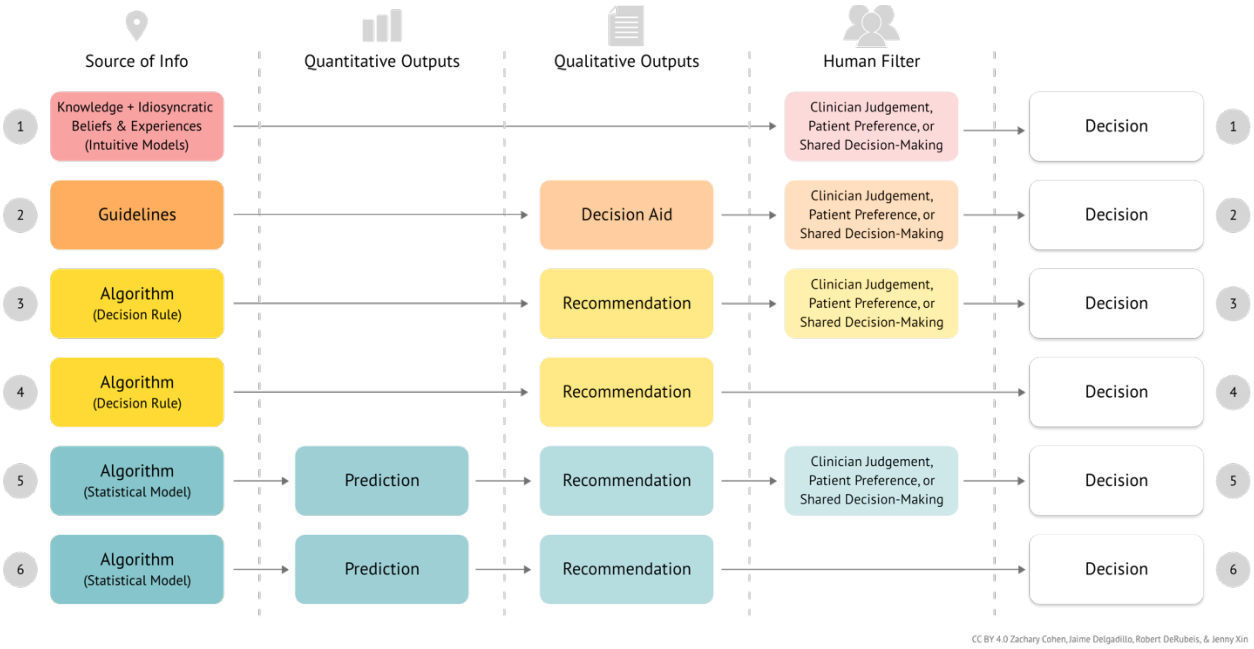

Figure 19.3. The Information-to-Decision Translational Framework

The three conceptual frameworks illustrated in Figures 19.1–19.3 can be used together to describe, organize, and critique a variety of approaches to personalization in the psychotherapy literature. However, none of these frameworks address the "How well does it work?" question. It is to this important issue —evaluation— that we now turn.

Evaluating Treatment Selection Approaches in Real-World Settings

Having already reviewed methods for quantifying prediction model accuracy, we will now focus on the more holistic evaluation of personalization *systems*. Approaches that provide inferential statistics about the expected performance of data-driven treatment selection rules are an active area of research (Kapelner et al., 2021; Luedtke & van der Laan, 2016a; Luedtke & van der Laan, 2017). Comparisons of outcomes for patients assigned to their model-predicted optimal treatment versus those assigned to their nonoptimal treatment, or those assigned randomly, can indicate whether there is signal in the set of predictions that rises above the noise. However, the effect size associated with the first comparison (optimal vs. nonoptimal) is unlikely to reflect any real-world scenario, given that it is difficult to imagine a clinician intentionally allocating all of their patients to their contra-indicated treatment. And although the comparison of optimal vs. random is reasonable and informative, it still might provide a biased estimate of the potential real-world impact of a personalization approach, insofar as current practice (e.g., allocation-as-usual) deviates from randomness (or quasi-

randomness). A more clinically relevant question is whether the recommendations derived from the model outperform typical or state-of-the-art methods that are in place for the decision of interest.

One way to address this question is through the collection of relevant baseline data (e.g., patient preference or clinician judgment) for secondary data analysis in the context of prospective randomized clinical trials. For example, a recent study by van Bronswijk, Lemmens et al. (2020) used an RCT of cognitive therapy vs. interpersonal therapy to compare model-based treatment recommendations against therapists' recommendations (recorded prior to randomization), which were based on case presentations of diagnostic workups. Model-based recommendations outperformed clinicians' recommendations. The authors also found evidence that patients who were randomized to the treatment recommended by the clinicians had worse outcomes than those randomized to the treatment identified by clinicians as nonpreferred, providing further caution against relying on intuitive treatment selection approaches.

Clinically relevant tests of model-driven personalization regimes are provided by prospective trials in which patients are randomized to receive treatment personalized based on the output of an algorithm (#6, Figure 19.3) or to treatment determined by means other than the application of the algorithm, such as *patient preference*, *clinician judgment*, or s*hared decision-making* (Figure 19.3, process #1), which is often described as *allocation-as-usual*.

It is understood that when an algorithm is disseminated, those using it will often allow input, after the fact, from the patient and clinician, in ways that could either enhance or detract from its utility. However, a test of a data-driven algorithm can be confounded if its application allows for a reversal of its recommendation by the clinician or patient. Likewise, the comparison condition can be contaminated if clinicians or clients in that condition are provided with information derived from the model. Further, differences in hope or expectancy can arise when the patient or clinician thinks they are working together on a treatment plan that has been "personalized" for that patient. Alternatively, they might be suspicious of a process whereby treatment is automatically selected by the algorithm without their input. Methods that can control for these potential confounds include incorporating patient preference as a feature in the statistical model or investigating separately the outcomes of patients whose recommended personalization was reversed by clinician judgment. Regarding personalization placebo effects, participants could be blinded as to whether they had been randomized to the personalized condition, or a double-blind could be implemented by providing clinicians in the control condition with fake (e.g., random) treatment recommendations.

If an unconfounded application of randomization is not possible (#6, Figure 19.3), an alternative experimental condition is for the recommendation from the

algorithm to inform the clinician's decision or a decision that is shared between the clinician and the client (#5, Figure 19.3). An advantage of this algorithm-informed selection process is that it reflects how algorithms are most likely to be implemented in practice. A limitation is that the utility, or accuracy, of the algorithm cannot be estimated independent of the input from the clinician and client. Relative to algorithm-based treatment selection, deviations from the algorithm's recommendation due to clinical judgment or patient influence could result in either improvements or worsening of patient outcomes. To the extent that a model's recommendation is not followed in the model-*informed* condition, the researcher cannot fully evaluate how well patients would have fared if their treatment had followed the model-derived plan. When departures from the model-derived recommendation are allowed, the number of reversals needs to be documented in order to elucidate the extent to which the experimental condition has been contaminated, and to permit secondary analyses of patients for whom the recommendations were and were not followed.

Evaluating Approaches to Tailoring Psychotherapy

When evaluating personalized psychotherapy that takes the form of selecting treatment components (3DP:L3) or adapting delivery style (3DP:L4), there are also several important methodological choices that must be made in relation to the control or comparison conditions. In order to address the specific question of the effectiveness of the personalization component itself, the personalized psychotherapy could be compared against a version of the same psychotherapy that lacks personalization, contains pseudo-personalization, uses randomized personalization, or that uses a comparable real-world approach to personalization (e.g., comparing data-driven selection of treatment components against patient preference-based selection of treatment components). The methods of blinding (if any) the researcher decides to employ in any of the above comparisons can have a significant impact on the interpretation of the study results. For any of the randomized comparisons discussed in this section, randomization could occur at the level of the patient, the clinician, or at the level of a clinic or hospital.

More complex trial designs can allow for compelling evaluations of different aspects of personalized psychotherapy. A study by Schulte et al. (1992) provides perhaps the best example of how the *personalization* in a psychotherapy personalization system can be evaluated. Phobic patients ($N = 120$) were randomly assigned to one of three conditions: (1) standardized *in vivo* exposure therapy, (2) personalized treatment in which the therapist used their case conceptualization to design a specific treatment plan for their patient that could use all therapeutic methods commonly employed in behavior therapy and cognitive therapy, or (3) a yoked control group, in which each patient's treatment followed a manual that was

created based on the treatment records of a specific patient in the personalized treatment condition with whom they were paired. A highly clinically relevant question – "Does this personalized therapy outperform standard evidence-based treatment?"– can be answered through the comparison of the standardized and personalized conditions. This contrast, however, does not allow for a pure evaluation of the personalization itself, due to differences in technique use (e.g., those in the standardized condition received more in vivo exposure and less cognitive methods) and treatment duration (patients in the personalized therapy received, on average, six more sessions) that arose due to the flexibility in the personalized therapy. The comparison of the yoked condition against personalized therapy controls for these differences in technique and duration, thus allowing for the evaluation of the more scientifically interesting question: "What is the specific value of the *personalization* in this individually tailored psychotherapy?" Schulte et al. (1992) found that standardized exposure therapy outperformed personalized therapy (comparison 1). They also found that the clinical decision-making that had guided the personalization had no impact on patient outcomes (comparison 2). To truly learn the value of psychotherapy personalization, more studies that allow for these nuanced comparisons (e.g., Kerns et al., 2014) are needed.

How Personalization Approaches Are Defined and Presented

Personalization systems need to be clearly defined, reproducible, and trainable. These requirements are similar to what is required for the evaluation of a structured psychotherapy, which would be unlikely to have a meaningful real-world impact in clinical settings if new practitioners could not be trained in its delivery. Similarly, warning flags should be raised if the description of the system through which personalization occurs is insufficiently detailed, or if the description reveals that the process would be unlikely to be replicable or trainable. Bonnett et al. (2019) discussed four different models of how clinical prediction models can be presented in order to serve as practical decision tools for clinicians. These models differ in complexity and sophistication; from simple point-scoring systems to determine if a patient's profile surpasses the threshold for a specific clinical decision, to fully computerized interfaces that automate complex statistical algorithms using a user-friendly input device. All of these options have advantages and disadvantages, and the choice of interface broadly depends on the clinical setting and the complexity of the underlying algorithm.

Berger et al. (2014) provide an example of a well-defined, replicable, data-informed approach to personalization. Patients with social anxiety, generalized anxiety, or panic were randomized to waitlist control, disorder-specific, or individually tailored internet-based guided self-help, where the intervention was tailored based on a series of pretreatment assessments (3DP:S4-L3-T1). Patients who were in the disorder-specific condition received only the content for their

principal diagnosis as determined by a diagnostic interview, whereas patients who were in the individually tailored condition received content specific to any problem area in which they scored at least 1.5 standard deviations higher than the normal population mean on the disorder-specific measure. In an example of a clearly defined data-informed approach, Delgadillo, Appleby et al. (2020) developed a data-driven decision algorithm called the Leeds Risk Index, which enables clinicians to easily implement a regression-based scoring system to decide which patients with common mental disorders should be offered low intensity guided self-help or high intensity psychotherapies. This method is highly standardized and guided by a simple visual diagram that fits with the definition of a "nomogram" as described by Bonnett et al. (2019).

Constructing Data-driven Personalized Treatment Decision Support Systems

Having already covered the ways in which information (i.e., a prediction or recommendation) from a treatment personalization approach could be used and evaluated, as well as the ways in which the translation of these outputs (regardless of their quality) can be influenced by the decision-making process, we now turn to the process of constructing the data-driven statistical models on which the most powerful treatment personalization approaches will be based. The various decisions that must be made throughout the stages of model development matter. Different researchers using the same data can draw very different conclusions about how to personalize treatment, as demonstrated in a recent "crowdsourced" analysis in which 12 research teams analyzed the same EMA dataset to select treatment targets (Bastiaansen et al., 2020). For historical perspectives on approaches to building statistical models for treatment selection in psychotherapy, see recent reviews by Cohen and DeRubeis (2018) and Cohen et al. (2019).

Sample Considerations

The limited sample sizes available in randomized clinical trials are often insufficient to adequately power the evaluation of prescriptive models. Luedtke and colleagues' (2019) simulations suggest that, in order to have adequate statistical power to develop and evaluate models and to detect clinically significant marginal improvements associated with treatment selection, at least 300 patients per treatment arm are needed. Lorenzo-Luaces et al. (2020) reviewed personalized medicine studies comparing CBT to other treatments for depression and identified 19 published multivariable prediction models, none of which met Luedtke and colleagues' sample size recommendation. As a solution to the sample size issues discussed throughout this chapter, Kessler and colleagues (2019) proposed that researchers first begin by developing models using a large observational (non-

randomized) datasets. Archival data will often lack the richness of datasets from RCTs, and such samples rarely provide the tight control over treatment parameters and assessment that is afforded by RCTs. Thus, the authors proposed an iterative process, whereby models developed in observational samples are subsequently examined in RCT datasets, and then tested prospectively in experimental studies. Delgadillo and Lutz (2020) described a similar development pathway where clinical prediction models are initially trained in large ($N > 1000$) observational datasets using an external cross-validation design, and then are prospectively tested in field-tests (e.g., small-scale prospective studies), randomized controlled trials, and then in implementation case studies that examine their effectiveness in routine care.

Supervised vs. Unsupervised Learning Approaches

One important initial decision when creating a data-driven personalization system is whether to apply a *supervised* or *unsupervised* modeling approach, the goals of which can differ. Supervised learning approaches aim to predict an outcome based on a number of inputs, whereas unsupervised approaches aim to model associations or patterns among a set of inputs, without the involvement of a specified outcome measure (Hastie et al., 2009).

A typical use-case for unsupervised approaches in treatment personalization is when information about patients is used to generate categories or dimensions, typically without the use of information about the treatment(s) received or outcomes observed during treatment. Associations with outcomes are examined only after the data-driven categories or dimensions are created. Common types of unsupervised approaches involve *clustering* and *dimensionality reduction*. Dimensionality reduction approaches include principal components analysis and exploratory or confirmatory factor analysis. Worked examples of clustering approaches that can inform psychotherapy treatment selection can be seen for latent profile analysis (Saunders et al., 2016), latent variable mixture modeling (Saunders et al., 2020), and agglomerative hierarchical and *k*-means cluster analysis (Deckersbach et al., 2016).

A wide variety of supervised learning approaches have been applied in mental health predictive modeling (see the previous section), ranging from classic ordinary least squares linear regression to non-parametric machine learning approaches, including random forests and Bayesian additive regression trees. The majority of the data-driven efforts we have described in this chapter aim to take advantage of associations between patient characteristics that can be ascertained prior to selecting a treatment (3DP:T1) and the primary outcomes resulting from the treatments under consideration (e.g., symptom reduction). Supervised learning approaches can also be employed to predict secondary outcomes (e.g., treatment dropout; Lutz et al., 2018) that are presumed to be related to the outcome of

interest (e.g., symptom response). When selecting secondary outcomes as targets for supervised personalization models, it is sensible to leverage presumed mediators of change. For example, in face-to-face therapy, common targets include attendance (e.g., Davis et al., 2020), engagement with the procedures of the therapy, homework compliance, or measures of the therapeutic alliance (e.g., Rubel et al., 2020). If there is a strong link between a presumed mediator and the outcome of interest, a model that predicts patients' scores on the mediator will tend also to predict outcome.

Prognostic and Prescriptive Modeling

The PROGRESS Group defined a prognostic factor as "any measure that, among people with a given health condition (that is, a start point), is associated with a subsequent clinical outcome (an endpoint)" (Hemingway et al., 2013). Models comprising prognostic variables (Steyerberg et al., 2013) can be contrasted with those that include prescriptive variables (also known as moderators or effect modifiers), which affect the direction or strength of the differences in outcome between two or more treatments (Baron & Kenny, 1986). A prescriptive variable is one that interacts with the treatment variable to predict differential response between treatments, and thus is informative for treatment selection models. See Cohen et al. (2019) for a comprehensive discussion of prognostic and prescriptive variables. For the purpose of this chapter, we will define prognostic models as any models that are constructed in a way that aims to predict an outcome of interest for a specific population, but without including treatment-moderator interaction terms. For example, a model that is constructed using only data from clients with depression who have been treated with psychodynamic therapy would be prognostic, because there is no way to know whether the predictive relationships instantiated in the model are identical to or different from those that would be observed had the same or similar individuals from the same population been treated with CBT, or medication, or had they received no treatment at all. Similarly, a model constructed using a mixed sample of individuals with depression, some of whom were treated with psychodynamic therapy, others of whom were treated with medication, would be classified as prognostic if information about *which* treatment each individual received was not available to the model, because there would be no explicit mechanism to capture differential response. Although this chapter is mostly focused on prognostic models of treatment response, another relevant form of prognostic modeling is risk-prediction (e.g., Ebert et al., 2019), which can be used to identify individuals at risk for developing problems or disorders (e.g., Kiekens et al., 2019; King et al., 2008).

The distinction between prognostic and prescriptive clinical prediction models takes on special importance in the arena of personalization. A systematic

review of methods for regression-based clinical prediction modeling by Rekkas and colleagues (2019) described three categories: risk-based methods (prognostic factors only), optimal treatment regime methods (treatment effect modifiers or prescriptive factors), and treatment effect modeling methods (which can include both prognostic and prescriptive factors). The goal of a treatment personalization approach (see Figure 19.1, *Level of Intervention* and *Timing* dimensions) can inform the decision of which modeling technique to use. For example, efforts to ensure that clients are initially triaged to the appropriate *level* of care are often amenable to prognostic modeling approaches (discussed in the previous section; see also Cohen & DeRubeis, 2018), whereas efforts to select treatment packages are often well-served by prescriptive modeling approaches.

Variable Selection

An important step of the *feature engineering* phase of constructing data-driven statistical models is the selection of *which* variables to include in the model. Different approaches applied to the same dataset can generate conflicting information about variable importance (Cohen et al., 2020). Repeated failures to identify any single predictive variable with large enough effects to guide treatment personalization (Cuijpers, Karyotaki, Reijnders, & Huibers, 2018; Simon & Perlis, 2010) reflect the need for multivariable approaches that can take into account the cumulative signal from multiple predictive factors (see Cohen & DeRubeis [2018] for a discussion of this issue). The selection process can itself have multiple stages, which can be guided by practical, theoretical, or empirical considerations. Methods have been developed specifically for selecting variables in the context of treatment personalization (Bossuyt & Parvin, 2015; Janes et al., 2011). As part of the STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative, Sauerbrei and colleagues (2020) published an extensive review of variable selection in which they summarize strengths and weaknesses of different approaches. They conclude that there is insufficient evidence to strongly recommend any single approach to variable selection, and that more research comparing different techniques is needed. Feature selection should incorporate cross-validation procedures that aim to select variables that are more likely to generalize to new samples and that are not overly influenced by overfitting (Rodriguez et al., 2009).

Machine Learning

Although the exact definition of machine learning remains elusive, this family of modeling approaches comprises methods that learn automatically from data, often flexibly capturing nonlinear associations and higher-order interactions. Uses of statistical modeling to draw conclusions from data can be broadly

described as either serving the goal of explanation/inference (which seeks to extract information about the nature of relations between variables) or prediction (Shmueli, 2010). Breiman (2001) described these, respectively, as "The Data Modeling Culture" and "The Algorithmic Modeling Culture." The latter approach is most often associated with machine learning. Arnold and colleagues (2020) discuss the implications of the differences between prediction and "causal explanation" on the use of machine learning methods. As a comprehensive discussion of machine learning is beyond the scope of this chapter, we refer the interested reader to several recent reviews, which can provide: a broad background on machine learning in clinical psychology (Coutanche & Hallion, 2019; Dwyer et al., 2018); a more quantitative review of machine learning-based prediction for mental health outcomes (Salazar de Pablo et al., 2020); and a review specifically focused on machine learning in psychotherapy research (Aafjes-van Doorn et al., 2020).

Machine-learning algorithms have gained widespread popularity in the mental health treatment outcome prediction literature (Lee et al., 2018). Novel modeling approaches have been developed that focus specifically on interactions and are thus well-suited to treatment selection (e.g., qualitative interaction trees; Dusseldorp & Van Mechelen, 2014). See Kapelner et al. (2021) for an extensive discussion of recent developments in statistical modeling. In mental health prediction, several studies have reported superior predictive performance of machine learning over classic regression (Kessler et al., 2016; Rosellini et al., 2018; Wardenaar et al., 2014; Webb, Cohen et al., 2019; c.f., Symons et al., 2020). Vollmer and colleagues (2020) set forth a helpful series of 20 critical questions (focused on six areas: inception, study, statistical methods, reproducibility, impact evaluation, and implementation) for all stakeholders involved in health-focused machine-learning efforts that aim to "provide a framework for research groups to inform the design, conduct, and reporting; for editors and peer reviewers to evaluate contributions to the literature; and for patients, clinicians, and policy makers to critically appraise where new findings may deliver patient benefit" (Vollmer et al., 2020). Although machine-learning approaches like neural networks, deep learning, boosting, support vector machines, and random forests – collectively called pure prediction algorithms by Efron (2020) – can result in improved predictive performance, key questions remain about when and how they can best be implemented (see the limitations section for further discussion).

Selecting a specific model or modeling approach to use can be challenging, especially given the abundance of publications in this space (Adibi et al., 2020). One way to inform this decision is to construct multiple predictive models in a sample using different approaches and then compare their performance using cross-validation, bootstrapping, or a held-out test sample (e.g., Delgadillo, Rubel, & Barkham, 2020; Hilbert et al., 2020; Webb, Cohen et al., 2019). A more advanced approach is to develop an ensemble in which multiple models are

combined to optimize performance. An example of an ensemble machine-learning approach is "super learning" (van der Laan et al., 2007), which uses cross-validation to select weights for a set of algorithms, and can also be used to construct ensembles for dynamic treatment regimes (Luedtke & van der Laan, 2016b). After constructing the ensemble approach, the fit of the individual models can be compared against the ensemble; for two examples using the SuperLearner package in R (Polley et al., 2016), see Rosellini et al. (2018) and Webb, Cohen et al. (2019).

Novel Research Designs for Constructing Data-Driven Personalized Treatments

Novel trial designs such as balanced fractional factorial trials (e.g., Watkins et al., 2016) and sequential multiple assignment randomized trials (SMART; e.g., Southward & Sauer-Zavala, 2020) are promising complements to more data-analytic ideographic (e.g., Altman et al., 2020) and nomothetic (e.g., Pearson et al., 2019) approaches to understanding the links between specific psychotherapy techniques and outcomes, and individual differences in response to different therapeutic elements. Clinical trials based on these new paradigms have the potential to help overcome some of the main barriers to understanding how and for whom treatments work, such as the sample size constraints that hampered the field's ability to draw meaningful conclusions from the dismantling or component trial literature (Cuijpers et al., 2017). In the context of digital therapeutics, these highly efficient experimental designs can facilitate the rapid development and evaluation of scalable personalized treatments (Uwatoko et al., 2018).

Challenges: Analytic, Methodological, and Statistical Problems with the Existing Literature on Personalized Psychotherapy

Personalized and precision approaches have garnered significant enthusiasm in mental health (Hollon et al., 2019; Holmes et al., 2018) and medicine more broadly (Collins & Varmus, 2015). Unfortunately, many – but not all – of these approaches have failed to live up to their promise (Joyner & Paneth, 2019; Prasad et al., 2016; Senn, 2018; Wilkinson et al., 2020). If personalized and precision mental health are going to make meaningful impacts on real-world

patient care, the methodological rigor of the development, reporting, and evaluation of these approaches must be improved.

Defining and Reporting Personalization Systems

Sufficient detail must be provided to allow a proper evaluation of a system for personalizing treatment. Multiple reviews suggest that this essential criterion is not being met. Using the TRIPOD checklist, Heus and colleagues (2018) reviewed publications reporting multivariable prediction models in top journals across 37 clinical domains. They identified 170 models and found that over 80% of the models were lacking key information, such as model specification or performance. These findings are reflected in the psychotherapy personalization literature as well. For example, consider Johansson et al.'s (2012) description of a personalization approach from their small RCT ($N = 40$ vs. 39) comparing standardized vs. individually tailored digital CBT: "An individualized treatment plan was prepared for each participant randomized to the tailored treatment. The treatment plans were formed by discussion in the research group and were mainly based on the SCID interview and results from self-report measures" (Johansson et al., 2012, p. 5). Not only is this description of the personalization insufficiently detailed to allow a reader to implement or evaluate the approach, but it also reveals levels of clinical judgement and flexibility that greatly decrease the likelihood that the approach could be replicated or disseminated. Similar issues regarding detail or flexibility can be seen across the literature on personalized therapy (Berger et al., 2017; Carlbring et al., 2011; Evers et al., 2014; Kraepelien et al., 2018, 2019; van Beugen et al., 2016).

Current Limitations of Machine Learning

The marriage of machine learning and precision medicine has also come under recent scrutiny (Wilkinson et al., 2020). Substantial evidence suggests that machine learning approaches should not be assumed to be a panacea, and that their use should be justified through comparisons against more simple methods of constructing clinical prediction models. A multitude of primary analyses as well as meta-analyses have reported similar performance for logistic regression and machine learning in prediction across a variety of domains including: diabetes (Lynam et al., 2020), chronic diseases (Nusinovici et al., 2020), traumatic brain injury (Gravesteijn et al., 2020), heart failure (Desai et al., 2020), and psychological treatments for depression and anxiety (Bone et al., 2021). Some studies have found evidence for improvements in accuracy when using machine learning approaches (relative to logistic regression) with more complex data (e.g., rich EMR-based information; Desai et al., 2020). One complicating feature of the literature comparing machine-learning to non-machine-learning approaches is the

inconsistency with which several of the most commonly used modeling techniques are classified. For example, Dwyer et al. (2018) describe penalized regression approaches (such as lasso, ridge, and elastic net regression) as machine learning. However, a recent systematic review of clinical prediction models (Christodoulou et al., 2019) that reported no advantage of machine learning over logistic regression categorized penalized regressions under the umbrella of logistic regression. Instead of relying on blanket statements about the benefits (or lack thereof) of machine learning, contextualized comparisons of specific approaches should be pursued.

Evaluation

Rigorous evaluation of personalized treatment approaches and the clinical prediction models on which they are sometimes based is essential (Hingorani et al., 2013; Pencina et al., 2020). Accordingly, it is alarming that, regarding the proliferation of clinical prediction models across the medical literature, including mental health, only a small proportion have been subjected to external validation (Collins et al., 2014; Siontis et al., 2015). For example, a systematic review of clinical prediction models in patients with chronic obstructive pulmonary disease identified 408 published models, less than 10% of which had been externally validated (Bellou et al., 2019). Low frequency of external validation has also been noted in systematic reviews of clinical prediction models in breast cancer (29%; Phung et al., 2019), cardiovascular disease risk (36%; Damen et al., 2016) and outcomes (17%; Wessler et al., 2015), preeclampsia (6%; De Kat et al., 2019), contrast-induced acute kidney injury (11%; Allen et al., 2017), and colorectal cancer (He et al., 2019). Wynants and colleagues' (2020) "living systematic review" (Macdonald et al., 2020) identified 145 prediction models for COVID-19, and concluded that "proposed models are poorly reported, at high risk of bias, and their reported performance is probably optimistic. Hence, we do not recommend any of these reported prediction models for use in current practice" (Wynants et al., 2020, p. 2). The importance of proper external validation can be seen in Gupta and colleagues' (2020) prospective evaluation of 22 published prediction models for COVID-19. They found, using decision curve analysis, that not a single multivariable model evidenced consistently higher net benefit for clinical deterioration or mortality than applying a rubric based on a single indicator (i.e., admission oxygen saturation on room air as a predictor of clinical deterioration; patient age for mortality) to stratify treatment, across a range of threshold probabilities. In summary, although well-established clinical prediction models in medicine do exist (e.g., Han et al., 2014; Hippisley-Cox et al., 2008; Maeda et al., 2019), the vast majority have either been poorly constructed, inadequately described, or insufficiently validated, and when evaluated, their clinical utility has often been negligible.

A recent systematic review and meta-analysis of precision psychiatry identified 584 studies in which prediction models (diagnostic, prognostic, or prescriptive) had been developed, only 10.4% and 4.6% of which reported having conducted proper internal and external validation, respectively (Salazar de Pablo et al., 2020). External cross-validation of the performance of predictive treatment selection models in psychotherapy has been exceedingly rare, perhaps in part due to the lack of overlap between variables across studies. One of the only attempts at cross-trial prediction of psychotherapy outcomes that used randomized data revealed significant reductions in the estimated advantage of treatment selection: van Bronswijk, Bruijniks et al. (2020) developed a model in an RCT comparing CBT and IPT, and found that the effect size estimate (generated via internal cross-validation) of receiving the model-indicated treatment shrank by 77% when that model was used to predict outcomes in a dataset from a second RCT of the same two treatments. However, there are examples of studies using naturalistic data that provide evidence for the generalizability of clinical prediction models in external test samples (e.g., Delgadillo & Gonzalez Salas Duhne, 2020; Delgadillo et al., 2017; Schwartz et al., 2020), some of which report impressive prediction accuracy in external samples from different psychological services that were not used to train the prediction model (Bone et al., 2021).

Much of the published literature on the prospective evaluation of personalized approaches to psychological interventions involves clinical trials that were designed to test the effectiveness of the intervention package and not the specific benefit of personalization. Myriad trials have compared personalized or tailored approaches to usual-care or control conditions (e.g., Buhrman et al., 2015; Carlbring et al., 2011; Chorpita et al., 2017; Day et al., 2013; Hallgren et al., 2015; Karyotaki et al., 2019; Nordgren et al., 2014; Silfvernagel et al., 2012, 2018; Strandskov et al., 2017; Twomey et al., 2017; van Beugen et al., 2016; Weisel et al., 2019). However, because these studies lacked analogous non-personalized comparisons, there is no way to know how much, if any, of the advantages for the personalized interventions (for those that reported them) was due to the fact that they were tailored to the patients. These trials, while useful for understanding whether the interventions themselves are more effective than the conditions with which they are compared, fail to provide evidence for the *specific* value of personalization. When a personalized psychotherapy is compared to a non-personalized variant, reduced effect sizes for specific tests of personalization should be expected (relative to comparisons against control comparisons); thus, larger samples will be required to adequately power these trials.

When the specific effects of personalization have been properly evaluated in randomized comparisons of personalized treatments against equivalent standard or static interventions (such tests are rare to begin with), the results have been mixed. Such comparisons, when based on non-data-driven approaches, have often found that treatment personalization failed to improve outcomes (e.g., Batterham

et al., 2018; Johansson et al., 2012; Kerns et al., 2014; Moritz et al., 2016) (cf. Cheavens et al., 2012; Miller et al., 2005).

A Cautionary Example from Personalized Pharmacotherapy

Recent findings from the pharmacotherapy research literature also underscore the importance of evaluating personalized treatment approaches. Even before the recent focus on precision psychiatry (e.g., pharmacogenetic and biomarker approaches to medication selection), guidelines existed describing how to personalize treatment to a patient. For example, when an inadequate response was observed to a given antidepressant, changing the dosage of the medication, augmenting the medication with a second one, or switching to a different class of antidepressants were (and still are) widely accepted "best practices." As we have seen in other contexts, *best* does not necessarily mean "validated," or even "tested." In comparisons of the effects of switching classes versus switching within the same class of antidepressants, some have found small or inconsistent increments of response or remission (Papakostas & Fava, 2008), and others have reported no advantage of switching at all (Bschor & Baethge, 2010; Rush et al., 2006; Souery et al., 2011). In regard to dosage adjustment as a way to optimize treatment response, a meta-analysis by Furukawa, Salanti and colleagues (2020) of 123 RCTs comprising 30K participants found no evidence of efficacy benefit of dose adjustment beyond the minimum licensed dose for SSRIs, venlafaxine, or mirtazapine. That these widely used and profession-endorsed methods of personalizing pharmacotherapy (Bauer et al., 2007) lack consistent supportive evidence of efficacy (c.f., Trivedi et al., 2004) highlights the importance of rigorous evaluation of personalized treatment approaches in mental health, including in psychotherapy.

Conclusions

An Evolutionary Perspective

Over 50 years have passed since Gordon Paul's call for more personalized approaches to psychotherapy, based on an understanding of *what works for whom* (Paul, 1967). Since then, numerous approaches to personalization have emerged. In this Chapter, we have attempted to examine these approaches from a historical and conceptual perspective. We have described this literature as gradually evolving from idiosyncratic approaches that are informed by clinical intuition towards algorithmic approaches that are informed by empirical data. Evolution is characterized by the diversification of species, by haphazard mutations in their

traits, and by the persistence or decline of traits through natural selection. As such, we have observed a similar process along the history of personalized psychotherapy: intuitions bred theories; theories bred data; and data gave rise to statistical algorithms (Figure 19.2). Today, the field of psychotherapy finds itself crossing paths with the fields of big data, predictive analytics, machine learning, and artificial intelligence. Along this evolutionary pathway we find a richly diverse ecosystem of personalization. Some approaches have been highly influential and persistent, such as the disorder-specific treatment selection approach that is widely adopted in contemporary clinical guidelines for evidence-based treatment. Others have failed to become mainstream practice or have fallen out of favor over time. This can be seen as a form of natural selection in the field of psychotherapy personalization. We have offered a taxonomy of features that can help characterize these diverse approaches, along the dimensions of *structure*, *level of intervention* and *time* (see Figure 19.1; 3DP: The Three Dimensions of Personalization). More recent approaches are characterized by greater complexity and sophistication, leveraging advances in methodology and data science. Despite this progress, we cannot yet confidently assert that recent data-driven approaches to personalization are more clinically effective. The evidence on data-driven approaches is still embryonic and we do not yet have the clinical trials or meta-analytic evidence needed to clearly demonstrate their advantage over standardized (e.g., generic, not explicitly personalized) psychotherapy. Taking a historical view, formalized approaches to personalization in psychotherapy are still in their early days. Despite the numerous studies cited in this chapter, the status of the evidence-base as a whole should be viewed as preliminary, with more research needed prior to clinical implications being derived with confidence. Nevertheless, data-driven approaches are emerging at a rapid pace and are likely to lead to future breakthroughs in our ability to improve the effectiveness of psychological care.

Current Challenges and Future Directions

The progress of this field, to some extent, will depend on a concerted effort by psychotherapy researchers, in collaboration with their colleagues in statistics and other areas, to overcome some of the methodological limitations that we have outlined in this chapter. Furthermore, the advent of data-driven decision tools also raises important questions related to the ethical implementation of such technologies in routine care.

Methodological Issues

In our view, researchers' enthusiasm for new technologies (e.g., new data collection techniques such as ecological momentary assessments; computerized treatment modules; network modeling; machine learning analyses) has often taken

precedence over basic aspects of methodological rigor. Methodological issues that have been neglected in this literature include the following:

- A lack of attention to statistical power considerations, exemplified by the development of clinical prediction models in inadequately small samples.
- A lack of attention to the problem of overfitting, exemplified by statistical models that are tested within the same sample used to construct the model without protective measures such as cross-validation, resampling or regularization techniques.
- A lack of attention to generalizability issues, exemplified by the predominance of prediction and personalization algorithms that are tested without the use of an external test sample.
- Poor specification and description of personalization models (e.g., their development and implementation), which limits the possibility of independent replication by other independent scientists.
- A lack of preregistration and other open-science best-practices in secondary data-analytic personalization efforts.
- The development of overly complex "black box" models, exemplified by complicated treatment selection and prediction methods that are not easily explainable to therapists or patients, thus potentially limiting their implementation in clinical practice.
- A scarcity of personalization models or tools that are designed using data from typical clinical populations, exemplified by models built on efficacy trials data which may not necessarily generalize to ordinary healthcare samples. Also lacking are prediction models constructed and evaluated using so-called non-WEIRD samples (western, educated, industrialized, rich and democratic).
- A scarcity of prospective tests of personalization methods, and particularly tests using rigorous experimental designs such as randomized clinical trials of personalization methods that are specified *a priori*.

Implementation issues

As reliable evidence for the value of personalized approaches accumulates, the field will need to turn to issues of dissemination and implementation. Some relevant questions in this regard are the following:

- How common is bias in data-driven algorithms? What are the causes of biased algorithms and what can be done to detect and mitigate their impact? The consequences of biased algorithms include predictions that are less accurate for underrepresented minority populations, and unfair

resource allocation, as when certain groups are denied equitable access to treatments from which they could benefit (Obermeyer et al., 2019).

- Do clinicians trust data-driven algorithms? How can they be effectively trained to use such algorithms? Would they follow an algorithm when their clinical intuition conflicts with its recommendation? Are some clinicians more likely to use algorithms compared to others? Which clinical decisions are better left to algorithms and which are better left to clinical experience?
- How do patients feel about the use of algorithms to inform their treatment? Is it necessary for patients to understand how these algorithms work? What if the patient's preferences are at odds with the recommendations from an algorithm?
- What role can industry play in the future of personalized psychotherapy? How can artificial intelligence and decision support tools for psychotherapy be implemented at scale? What are the implications for confidentiality and data protection? What are the implications for data sharing and open science? What are the implications for the ethical governance and regulation of future artificial intelligence devices?
- What are the implications for the training of new generations of psychotherapists? What skills and knowledge will new psychotherapists need to acquire in order to understand and use data-driven personalization methods (Lutz et al., 2020)?

Recent evidence suggests that although progress is being made on some of the issues in reporting and statistical methodology described in this chapter (Najafabadi et al., 2020), significant work remains. Recent developments described in this chapter provide reasons to be optimistic. Prospective demonstrations of improved outcomes achieved through data-driven personalization have begun to emerge (Lutz et al., 2021; Delgadillo et al., under review). It is likely that, in the not-so-distant future, data-driven decision support tools will be part of the arsenal of tools available to psychotherapists.

References

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2020). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research, 31*(1), 92–116. https://doi.org/10.1080/10503307.2020.1808729

Adibi, A., Sadatsafavi, M., & Ioannidis, J. P. (2020). Validation and utility testing of clinical prediction models: Time to change the approach. *Journal of the American Medical Association, 324*(3), 234–236. https://doi.org/10.1001/jama.2020.1230

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., & Cohen, G. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*(3), 341–382. https://doi.org/10.1177/0011000005285875

Alba, A. C., Agoritsas, T., Walsh, M., Hanna, S., Iorio, A., Devereaux, P. J., McGinn, T., & Guyatt, G. (2017). Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. *Journal of the American Medical Association, 318*(14), 1377–1384. https://doi.org/10.1001/jama.2017.12126

Allen, D. W., Ma, B., Leung, K. C., Graham, M. M., Pannu, N., Traboulsi, M., Goodhart, D., Knudtson, M. L., & James, M. T. (2017). Risk prediction models for contrast-induced acute kidney injury accompanying cardiac catheterization: Systematic review and meta-analysis. *Canadian Journal of Cardiology*, *33*(6), 724–736. https://doi.org/10.1016/j.cjca.2017.01.018

Altman, A. D., Shapiro, L. A., & Fisher, A. J. (2020). Why does therapy work? An idiographic approach to explore mechanisms of change over the course of psychotherapy using digital assessments. *Frontiers in Psychology*, *11*, 782. https://doi.org/10.3389/fpsyg.2020.00782

Antezana, G., Bidargaddi, N., Blake, V., Schrader, G., Kaambwa, B., Quinn, S., Orlowski, S., Winsall, M., & Battersby, M. (2015). Development of an online well-being intervention for young people: An evaluation protocol. *JMIR Research Protocols, 4*(2), e48. https://doi.org/10.2196/resprot.4098

Archer, L., Snell, K. I. E., Ensor, J., Hudda, M. T., Collins, G. S., & Riley, R. D. (2020). Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Statistics in Medicine. 40*(1), 133–146. https://doi.org/10.1002/sim.8766

Arnold, K. F., Davies, V., de Kamps, M., Tennant, P. W., Mbotwa, J., & Gilthorpe, M. S. (2020). Reflections on modern methods: Generalized linear models for prognosis and intervention—Theory, practice and implications for machine learning. *International Journal of Epidemiology*, dyaa049. https://doi.org/10.1093/ije/dyaa049

Arthur, A. R. (2001). Personality, epistemology and psychotherapists' choice of theoretical model: A review and analysis. *European Journal of Psychotherapy, Counselling & Health*, *4*(1), 45–64. https://doi.org/10.1080/13642530110040082

Bae, S., Chung, T., Ferreira, D., Dey, A. K., & Suffoletto, B. (2018). Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive Behaviors*, *83*, 42–47. https://doi.org/10.1016/j.addbeh.2017.11.039

Barlow, D. H., Farchione, T. J., Bullis, J. R., Gallagher, M. W., Murray-Latin, H., Sauer-Zavala, S., Bentley, K. H., Thompson-Hollands, J., Conklin, L. R., Boswell, J. F., Ametaj, A., Carl, J. R., Boettcher, H. T., & Cassiello-Robbins, C. (2017). The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: A randomized clinical trial. *JAMA Psychiatry, 74*(9), 875–884. https://doi.org/10.1001/jamapsychiatry.2017.2164

Barlow, D. H., Farchione, T. J., Fairholme, C., Ellard, K., Boisseau, C., Allen, L., & Ehrenreich-May, J. (2011). *The unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. Oxford University Press.

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*(6), 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F. J., Boker, S. M., Ceulemans, E., Chen, M., Chow, S. M., de Jonge, P., Emerencia, A. C., Epskamp, S., Fisher, A. J., Hamaker, E. L., Kuppens, P., Lutz, W., Meyer, M. J., Moulder, R., Oravecz, Z., Riese, H., Rubel, J., Ryan, … Bringmann, L. F. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal of Psychosomatic Research*, *137*, 110211. https://doi.org/10.1016/j.jpsychores.2020.110211

Batterham, P. J., Calear, A. L., Farrer, L., McCallum, S. M., & Cheng, V. W. S. (2018). FitMindKit: Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. *Internet Interventions*, *12*, 91–99. https://doi.org/10.1016/j.invent.2017.08.002

Bauer, M., Bschor, T., Pfennig, A., Whybrow, P. C., Angst, J., Versiani, M., Möller, H.-J., Disorders, W. T. F. o. U. D., Bauer, M., & Bschor, T. (2007). World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders in primary care. *The World Journal of Biological Psychiatry*, *8*(2), 67–104. https://doi.org/10.1080/15622970701227829

Bell, I., & Mellor, D. (2009). Clinical judgements: Research and practice. *Australian Psychologist*, *44*(2), 112–121. https://doi.org/10.1080/00050060802550023

Bellou, V., Belbasis, L., Konstantinidis, A. K., Tzoulaki, I., & Evangelou, E. (2019). Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: Systematic review and critical appraisal. *British Medical Journal*, *367*, l5358. https://doi.org/10.1136/bmj.l5358

Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. *Journal of Counseling Psychology, 58*(3), 279–289. https://doi.org/10.1037/a0023626

Benner, P., & Tanner, C. (1987). How expert nurses use intuition. *The American Journal of Nursing*, *87*(1), 23–34.

Berger, T., Boettcher, J., & Caspar, F. (2014). Internet-based guided self-help for several anxiety disorders: A randomized controlled trial comparing a tailored with a standardized disorder-specific approach. *Psychotherapy*, *51*(2), 207–219. https://doi.org/10.1037/a0032527

Berger, T., Urech, A., Krieger, T., Stolz, T., Schulz, A., Vincent, A., Moser, C. T., Moritz, S., & Meyer, B. (2017). Effects of a transdiagnostic unguided Internet intervention ('velibra') for anxiety disorders in primary care: Results of a randomized controlled trial. *Psychological Medicine*, *47*(1), 67–80. https://doi.org/10.1017/S0033291716002270

Betan, E. J., & Binder, J. L. (2010). Clinical expertise in psychotherapy: How expert therapists use theory in generating case conceptualizations and interventions. *Journal of Contemporary Psychotherapy*, *40*(3), 141–152. https://doi.org/10.1007/s10879-010-9138-0

Beutler, L. E., Consoli, A. J., & Lane, G. (2005). Systematic treatment selection and prescriptive psychotherapy: An integrative eclectic approach. In J. C. Norcross, & M. R. Goldfried (Eds.), *Handbook of Psychotherapy Integration* (2nd ed., pp. 121–143).

Bidargaddi, N., Musiat, P., Winsall, M., Vogl, G., Blake, V., Quinn, S., Orlowski, S., Antezana, G., & Schrader, G. (2017). Efficacy of a web-based guided recommendation service for a curated list of readily available mental health and well-being mobile apps for young people: Randomized controlled trial. *Journal of Medical Internet Research*, *19*(5), e141. https://doi.org/10.2196/jmir.6775

Bone, C., Simmonds-Buckley, M., Thwaites, R., Sandford, D., Merzhvynska, M., Rubel, J., Deisenhofer, A-K., Lutz, W., & Delgadillo, J. (2021). Dynamic prediction of psychological treatment outcomes using routinely collected symptom data. *The Lancet Digital Health, 3*(4), e231-e240.

Bonnett, L. J., Snell, K. I., Collins, G. S., & Riley, R. D. (2019). Guide to presenting clinical prediction models for use in clinical settings. *British Medical Journal*, *365*, 1737. https://doi.org/10.1136/bmj.l737

Bossuyt, P. M., & Parvin, T. (2015). Evaluating biomarkers for guiding treatment decisions. *eJIFCC*, *26*(1), 63–70. https://www.ncbi.nlm.nih.gov/pubmed/27683482

Bowens, M., & Cooper, M. (2012). Development of a client feedback tool: A qualitative study of therapists' experiences of using the Therapy Personalisation Forms. *European Journal of Psychotherapy & Counselling*, *14*(1), 47–62. https://doi.org/10.1080/13642537.2012.652392

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://projecteuclid.org/euclid.ss/1009213726

Bschor, T., & Baethge, C. (2010). No evidence for switching the antidepressant: Systematic review and meta-analysis of RCTs of a common therapeutic strategy. *Acta Psychiatrica Scandinavica*, *121*(3), 174–179. https://doi.org/10.1111/j.1600-0447.2009.01458.x

Buhrman, M., Syk, M., Burvall, O., Hartig, T., Gordh, T., & Andersson, G. (2015). Individualized guided internet-delivered cognitive-behavior therapy for chronic pain patients with comorbid depression and anxiety. *The Clinical Journal of Pain*, *31*(6), 504–516. https://doi.org/10.1097/AJP.0000000000000176

Cabral, R. R., & Smith, T. B. (2011). Racial/ethnic matching of clients and therapists in mental health services: A meta-analytic review of preferences, perceptions, and outcomes. *Journal of Counseling Psychology*, *58*(4), 537–554. https://doi.org/10.1037/a0025266

Canadian Psychological Association. (2012). Evidence-based practice of psychological treatments: A Canadian perspective. *Canada: CPA Task Force on Evidence-Based Practice of Psychological Treatments*.

Captari, L. E., Hook, J. N., Hoyt, W., Davis, D. E., McElroy-Heltzel, S. E., & Worthington Jr, E. L. (2018). Integrating clients' religion and spirituality within psychotherapy: A comprehensive meta-analysis. *Journal of Clinical Psychology*, *74*(11), 1938–1951. https://doi.org/10.1002/jclp.22681

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. *Cognitive Behaviour Therapy*, *47*(1), 1–18. https://doi.org/10.1080/16506073.2017.1401115

Carlbring, P., Maurin, L., Torngren, C., Linna, E., Eriksson, T., Sparthan, E., Straat, M., Marquez von Hage, C., Bergman-Nordgren, L., & Andersson, G. (2011). Individually-tailored, internet-based treatment for anxiety disorders: A randomized controlled trial. *Behaviour Research and Therapy*, *49*(1), 18–24. https://doi.org/10.1016/j.brat.2010.10.002

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 7–18. https://doi.org/10.1037//0022-006x.66.1.7

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review Psychology*, *52*(1), 685–716. https://doi.org/10.1146/annurev.psych.52.1.685

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

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*, *77*(3), 566–579. https://doi.org/10.1037/a0014565

Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., Chiu, A. W., Letamendi, A. M., Tsai, K. H., & Krull, J. L. (2017). Child STEPs in California: A

cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology*, *85*(1), 13–25. https://doi.org/10.1037/ccp0000133

Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Modularity in the design and application of therapeutic interventions. *Applied and Preventive Psychology*, *11*(3), 141–156. https://doi.org/10.1016/j.appsy.2005.05.002

Chowdhary, N., Jotheeswaran, A. T., Nadkarni, A., Hollon, S. D., King, M., Jordans, M. J., Rahman, A., Verdeli, H., Araya, R., & Patel, V. (2014). The methods and outcomes of cultural adaptations of psychological treatments for depressive disorders: A systematic review. *Psychological Medicine*, *44*(6), 1131–1146. https://doi.org/10.1017/S0033291713001785

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

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. https://doi.org/10.1146/annurev-clinpsy-050817-084833

Cloitre, M., Cohen, Z. D., & Schnyder, U. (2020). Building a Science of Personalized Interventions for PTSD. In D. Forbes, J. I. Bisson, C. M. Monson, & L. Berliner (Eds.), *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies* (3rd ed., pp. 451–468). Guilford Publications.

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

Cohen, Z. D., Ashar, Y. K., & DeRubeis, D. J. (2019). Moving beyond "one size fits all". In S. Dimidjian (Ed.), *Evidence-based practice in action: Bridging clinical science and intervention* (pp. 111–132). Guilford Press.

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

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

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, *372*(9), 793–795. https://doi.org/10.1056/NEJMp1500523

Collins, G. S., de Groot, J. A., Dutton, S., Omar, O., Shanyinde, M., Tajar, A., Voysey, M., Wharton, R., Yu, L.-M., & Moons, K. G. (2014). External validation of multivariable

prediction models: A systematic review of methodological conduct and reporting. *BMC Medical Research Methodology*, *14*(1), 40. https://doi.org/10.1186/1471-2288-14-40

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). The TRIPOD Statement. *Circulation*, *131*(2), 211–219. https://doi.org//10.1161/CIRCULATIONAHA.114.014508

Cooper, M., & McLeod, J. (2011). Person-centered therapy: A pluralistic perspective. *Person-Centered & Experiential Psychotherapies, 10*(3), 210–223. https://doi.org/10.1080/14779757.2011.599517

Coull, G., & Morris, P. (2011). The clinical effectiveness of CBT-based guided self-help interventions for anxiety and depressive disorders: A systematic review. *Psychological Medicine, 41*(11), 2239–2252. https://doi.org/10.1017/S0033291711000900

Coutanche, M. N., & Hallion, L. S. (2019). Machine learning for clinical psychology and clinical neuroscience. In A. G. C. Wright & M. N. Hallquist (Eds.), *The Cambridge handbook of research methods in clinical psychology* (pp. 467–482). Cambridge University Press. https://doi.org/10.1017/9781316995808

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. https://doi.org/10.1037/h0043943

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Hollon, S. D. (2019). Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. *Psychotherapy Research*, *29*(1), 15–29. https://doi.org/10.1080/10503307.2017.1395922

Cuijpers, P., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2018). Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence. *Cognitive Behaviour Therapy*, *47*(2), 91–106. https://doi.org/10.1080/16506073.2017.1420098

Cuijpers, P., Karyotaki, E., Reijnders, M., Purgato, M., & Barbui, C. (2018). Psychotherapies for depression in low- and middle-income countries: A meta-analysis. *World Psychiatry*, *17*(1), 90–101. https://doi.org/10.1002/wps.20493

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

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*, *19*(1), 92–107. https://doi.org/10.1002/wps.20701

Dalgleish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of Consulting and Clinical Psychology*, *88*(3), 179–195. https://doi.org/10.1037/ccp0000482

Damen, J. A., Hooft, L., Schuit, E., Debray, T. P., Collins, G. S., Tzoulaki, I., Lassale, C. M., Siontis, G. C., Chiocchia, V., & Roberts, C. (2016). Prediction models for cardiovascular disease risk in the general population: Systematic review. *British Medical Journal*, *353,* i2416. https://doi.org/10.1136/bmj.i2416

Dance, K. A., & Neufeld, R. W. (1988). Aptitude-treatment interaction research in the clinical setting: A review of attempts to dispel the" patient uniformity" myth. *Psychological Bulletin*, *104*(2), 192–213. https://doi.org/10.1037/0033-2909.104.2.192

Davis, A., Smith, T., Talbot, J., Eldridge, C., & Betts, D. (2020). Predicting patient engagement in IAPT services: A statistical analysis of electronic health records. *Evidence Based Mental Health*, *23*(1), 8–14. https://doi.org/10.1136/ebmental-2019-300133

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674. https://doi.org/10.1126/science.2648573

Dawes, R. M., Faust, D., & Meehl, P. E. (1993). Statistical prediction versus clinical prediction: Improving what works. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 351–367). Lawrence Erlbaum Associates.

Day, V., McGrath, P. J., & Wojtowicz, M. (2013). Internet-based guided self-help for university students with anxiety, depression and stress: A randomized controlled clinical trial. *Behaviour Research and Therapy*, *51*(7), 344–351. https://doi.org/10.1016/j.brat.2013.03.003

De Kat, A. C., Hirst, J., Woodward, M., Kennedy, S., & Peters, S. A. (2019). Prediction models for preeclampsia: A systematic review. *Pregnancy Hypertension*, *16*, 48–66. https://doi.org/10.1016/j.preghy.2019.03.005

Deckersbach, T., Peters, A. T., Sylvia, L. G., Gold, A. K., da Silva Magalhaes, P. V., Henry, D. B., Frank, E., Otto, M. W., Berk, M., & Dougherty, D. D. (2016). A cluster analytic approach to identifying predictors and moderators of psychosocial treatment for bipolar depression: Results from STEP-BD. *Journal of Affective Disorders*, *203*, 152–157. https://doi.org/10.1016/j.jad.2016.03.064

Deisenhofer, A. K., Delgadillo, J., Rubel, J. A., Bohnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety*, *35*(6), 541–550. https://doi.org/10.1002/da.22755

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., Cohen, Z. D., DeRubeis, R. J., and Barkham, M. (*under review*). StratCare: A pragmatic, multi-site, single-blind, cluster randomised controlled trial of stratified care for depression.

Delgadillo, J., Appleby, S., Booth, S., Burnett, G., Carey, A., Edmeade, L., Green, S., Griffin, P., Johnson, E., Jones, R., Parker, P., Reeves-McLaren, L., & Lutz, W. (2020). The Leeds Risk Index: Field-Test of a stratified psychological treatment selection algorithm. *Psychotherapy and Psychosomatics*, *89*(3), 189–190. https://doi.org/10.1159/000505193

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

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

Delgadillo, J., & Lutz, W. (2020). A development pathway towards precision mental health care. *JAMA Psychiatry,* 77(9), 889–890. https://doi.org/10.1001/jamapsychiatry.2020.1048

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy*, *79*, 15–22. https://doi.org/10.1016/j.brat.2016.02.003

Delgadillo, J., Rubel, J., & Barkham, M. (2020). Towards personalized allocation of patients to therapists. *Journal of Consulting and Clinical Psychology*, *88*(9), 799–808. https://doi.org/10.1037/ccp0000507

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLOS One*, *9*(1), e83875. https://doi.org/10.1371/journal.pone.0083875

DeRubeis, R. J., 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*(3), 419–428. https://doi.org/10.1080/10503307.2013.838654

Desai, R. J., Wang, S. V., Vaduganathan, M., Evers, T., & Schneeweiss, S. (2020). Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. *JAMA Network Open*, *3*(1), e1918962-e1918962. https://doi.org/10.1001/jamanetworkopen.2019.18962

Dickens, C., Cherrington, A., Adeyemi, I., Roughley, K., Bower, P., Garrett, C., Bundy, C., & Coventry, P. (2013). Characteristics of psychological interventions that improve depression in people with coronary heart disease: A systematic review and meta-regression. *Psychosomatic Medicine*, *75*(2), 211–221. https://doi.org/10.1097/PSY.0b013e31827ac009

Dusseldorp, E., & Van Mechelen, I. (2014). Qualitative interaction trees: A tool to identify qualitative treatment-subgroup interactions. *Statistics in Medicine*, *33*(2), 219–237. https://doi.org/10.1002/sim.5933

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Ebert, D. D., Buntrock, C., Mortier, P., Auerbach, R., Weisel, K. K., Kessler, R. C., Cuijpers, P., Green, J. G., Kiekens, G., Nock, M. K., Demyttenaere, K., & Bruffaerts, R. (2019). Prediction of major depressive disorder onset in college students. *Depression and Anxiety*, *36*(4), 294–304. https://doi.org/10.1002/da.22867

Eells, T. D., Lombart, K. G., Kendjelic, E. M., Turner, L. C., & Lucas, C. P. (2005). The quality of psychotherapy case formulations: A comparison of expert, experienced, and novice cognitive-behavioral and psychodynamic therapists. *Journal of Consulting and Clinical Psychology*, *73*(4), 579–589. https://doi.org/10.1037/0022-006X.73.4.579

Efron, B. (2020). Prediction, Estimation, and Attribution. *Journal of the American Statistical Association*, *115*(530), 636–655. https://doi.org/10.1080/01621459.2020.1762613

Elwyn, G., O'Connor, A., Stacey, D., Volk, R., Edwards, A., Coulter, A., Thomson, R., Barratt, A., Barry, M., Bernstein, S., Butow, P., Clarke, A., Entwistle, V., Feldman-Stewart, D., Holmes-Rovner, M., Llewellyn-Thomas, H., Moumjid, N., Mulley, A., Ruland, C., … Whelan, T. (2006). Developing a quality criteria framework for patient decision aids: Online international Delphi consensus process. *British Medical Journal*, *333*(7565), 417. https://doi.org/10.1136/bmj.38926.629329.AE

Eskildsen, A., Reinholt, N., van Bronswijk, S., Brund, R. B. K., Christensen, A. B., Hvenegaard, M., Arendt, M., Alrø, A., Poulsen, S., Rosenberg, N. K., Huibers, M. J. H., & Arnfred, S. (2020). Personalized psychotherapy for outpatients with major depression and anxiety disorders: Transdiagnostic versus diagnosis-specific group cognitive behavioural therapy. *Cognitive Therapy and Research*, *44*(5), 988–1001. https://doi.org/10.1007/s10608-020-10116-1

Evers, A. W., Gieler, U., Hasenbring, M. I., & van Middendorp, H. (2014). Incorporating biopsychosocial characteristics into personalized healthcare: A clinical approach. *Psychotherapy and Psychosomatics*, *83*(3), 148–157. https://doi.org/10.1159/000358309

Fernandez, K. C., Fisher, A. J., & Chi, C. (2017). Development and initial implementation of the Dynamic Assessment Treatment Algorithm (DATA). *PLOS One*, *12*(6), e0178806. https://doi.org/10.1371/journal.pone.0178806

Field, M. J., & Lohr, K. N. (1990). Clinical practice guidelines. *Directions for a new program.* National Academy Press.

Finegan, M., Firth, N., Wojnarowski, C., & Delgadillo, J. (2018). Associations between socioeconomic status and psychological therapy outcomes: A systematic review and meta-analysis. *Depression and Anxiety*, *35*(6), 560–573. https://doi.org/10.1002/da.22765

Firth, N., Barkham, M., & Kellett, S. (2015). The clinical effectiveness of stepped care systems for depression in working age adults: A systematic review. *Journal of Affective Disorders*, *170*, 119–130. https://doi.org/10.1016/j.jad.2014.08.030

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

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

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, 115(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Foa, E. B., & Rothbaum, B. O. (1998). *Treatment manuals for practitioners*. Treating the trauma of rape: Cognitive-behavioral therapy for PTSD. Guilford Press. https://doi.org/10.1080/10503309912331332731

Forsell, E., Isacsson, N., Blom, K., Jernelöv, S., Ben Abdesslem, F., Lindefors, N., Boman, M., & Kaldo, V. (2019). Predicting treatment failure in regular care internet-delivered cognitive behavior therapy for depression and anxiety using only weekly symptom measures. *Journal of Consulting and Clinical Psychology, 88*(4), 311–321. https://doi.org/10.1037/ccp0000462

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

Friedl, N., Krieger, T., Chevreul, K., Hazo, J. B., Holtzmann, J., Hoogendoorn, M., Kleiboer, A., Mathiasen, K., Urech, A., & Riper, H. (2020). Using the Personalized Advantage Index for Individual Treatment Allocation to Blended Treatment or Treatment as Usual for Depression in Secondary Care. *Journal of clinical medicine*, *9*(2), 490.

Furukawa, T. A., Debray, T. P. A., Akechi, T., Yamada, M., Kato, T., Seo, M., & Efthimiou, O. (2020). Can personalized treatment prediction improve the outcomes, compared with the group average approach, in a randomized trial? Developing and validating a multivariable prediction model in a pragmatic megatrial of acute treatment for major depression. *Journal of Affective Disorders*, *274*, 690–697. https://doi.org/10.1016/j.jad.2020.05.141

Furukawa, T. A., Salanti, G., Cowen, P. J., Leucht, S., & Cipriani, A. (2020). No benefit from flexible titration above minimum licensed dose in prescribing antidepressants for major depression: Systematic review. *Acta Psychiatrica Scandinavica*, *141*(5), 401–409. https://doi.org/10.1111/acps.13145

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, *1*, 67–89. https://doi.org/10.1146/annurev.clinpsy.1.102803.143810

Garfield, S. L. (1996). Some problems associated with "validated" forms of psychotherapy. *Clinical psychology: Science and practice*, *3*(3), 218–229. https://doi.org/10.1111/j.1468-2850.1996.tb00073.x

Gaudiano, B. A., Brown, L. A., & Miller, I. W. (2011). Let your intuition be your guide? Individual differences in the evidence-based practice attitudes of psychotherapists. *Journal of Evaluation in Clinical Practice*, *17*(4), 628–634. https://doi.org/10.1111/j.1365-2753.2010.01508.x

Gelfand, L. A., Lorenzo-Luaces, L., & DeRubeis, R. J. (*under review*). Effects of patient response pattern distributions on comparisons of psychosocial treatments for major depressive disorder.

Gemmell, I., & Dunn, G. (2011). The statistical pitfalls of the partially randomized preference design in non-blinded trials of psychological interventions. *International Journal of Methods in Psychiatric Research*, *20*(1), 1–9. https://doi.org/10.1002/mpr.326

Goldstein, S. P., Evans, B. C., Flack, D., Juarascio, A., Manasse, S., Zhang, F., & Forman, E. M. (2017). Return of the JITAI: Applying a just-in-time adaptive intervention framework to the development of m-health solutions for addictive behaviors. *International Journal of Behavioral Medicine*, *24*(5), 673–682. https://doi.org/10.1007/s12529-016-9627-y

Gravesteijn, B. Y., Nieboer, D., Ercole, A., Lingsma, H. F., Nelson, D., Van Calster, B., Steyerberg, E. W., & collaborators, C.-T. (2020). Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. *Journal of Clinical Epidemiology*, *122*, 95–107. https://doi.org/10.1016/j.jclinepi.2020.03.005

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*(2), 293–323. https://doi.org/https://doi.org/10.1037/1076–8971.2.2.293

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*(1), 19–30. https://www.ncbi.nlm.nih.gov/pubmed/10752360

Gupta, R. K., Marks, M., Samuels, T. H. A., Luintel, A., Rampling, T., Chowdhury, H., Quartagno, M., Nair, A., Lipman, M., Abubakar, I., van Smeden, M., Wong, W. K., Williams, B., Noursadeghi, M., & Group, U. C.-R. (2020). Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: An observational cohort study. *European Respiratory Journal*, *56*, *2003498*. https://doi.org/10.1183/13993003.03498-2020

Hall, G. C., Ibaraki, A. Y., Huang, E. R., Marti, C. N., & Stice, E. (2016). A meta-analysis of cultural adaptations of psychological interventions. *Behavior Therapy*, *47*(6), 993–1014. https://doi.org/10.1016/j.beth.2016.09.005

Hallgren, M., Kraepelien, M., Lindefors, N., Zeebari, Z., Kaldo, V., & Forsell, Y. (2015). Physical exercise and internet-based cognitive–behavioural therapy in the treatment of depression: Randomised controlled trial. *The British Journal of Psychiatry*, *207*(3), 227–234. https://doi.org/10.1192/bjp.bp.114.160101

Han, J., King, N. K., Neilson, S. J., Gandhi, M. P., & Ng, I. (2014). External validation of the CRASH and IMPACT prognostic models in severe traumatic brain injury. *Journal of Neurotrauma*, *31*(13), 1146–1152. https://doi.org/10.1089/neu.2013.3003

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). Springer.

He, Y., Ong, Y., Li, X., Din, F. V., Brown, E., Timofeeva, M., Wang, Z., Farrington, S. M., Campbell, H., & Dunlop, M. G. (2019). Performance of prediction models on survival outcomes of colorectal cancer with surgical resection: A systematic review and meta-analysis. *Surgical Oncology*, *29*, 196–202. https://doi.org/10.1016/j.suronc.2019.05.014

Hemingway, H., Croft, P., Perel, P., Hayden, J. A., Abrams, K., Timmis, A., Briggs, A., Udumyan, R., Moons, K. G., Steyerberg, E. W., Roberts, I., Schroter, S., Altman, D. G., Riley, R. D., & Group, P. (2013). Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *British Medical Journal*, *346*, e5595. https://doi.org/10.1136/bmj.e5595

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Heus, P., Damen, J., Pajouheshnia, R., Scholten, R., Reitsma, J. B., Collins, G. S., Altman, D. G., Moons, K. G. M., & Hooft, L. (2018). Poor reporting of multivariable prediction model studies: Towards a targeted implementation strategy of the TRIPOD statement. *BMC Medicine*, *16*(1), 120. https://doi.org/10.1186/s12916-018-1099-2

Hilbert, K., Kunas, S. L., Lueken, U., Kathmann, N., Fydrich, T., & Fehm, L. (2020). Predicting cognitive behavioral therapy outcome in the outpatient sector based on clinical routine data: A machine learning approach. *Behaviour Research and Therapy*, *124*, 103530. https://doi.org/10.1016/j.brat.2019.103530

Hingorani, A. D., Windt, D. A., Riley, R. D., Abrams, K., Moons, K. G., Steyerberg, E. W., Schroter, S., Sauerbrei, W., Altman, D. G., Hemingway, H., & Group, P. (2013). Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *British Medical Journal*, *346*, e5793. https://doi.org/10.1136/bmj.e5793

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., Minhas, R., Sheikh, A., & Brindle, P. (2008). Predicting cardiovascular risk in England and Wales: Prospective derivation and validation of QRISK2. *British Medical Journal*, *336*(7659), 1475–1482. https://doi.org/10.1136/bmj.39609.449676.25

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention Sscience: Process-based therapy. *Clinical Psychological Science*, *7*(1), 37–50. https://doi.org/10.1177/2167702618772296

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

Holmes, E. A., Ghaderi, A., Harmer, C. J., Ramchandani, P. G., Cuijpers, P., Morrison, A. P., Roiser, J. P., Bockting, C. L. H., O'Connor, R. C., Shafran, R., Moulds, M. L., & Craske, M. G. (2018). The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *Lancet Psychiatry*, *5*(3), 237–286. https://doi.org/10.1016/S2215-0366(17)30513-8

Huang, Y., Li, W., Macheret, F., Gabriel, R. A., & Ohno-Machado, L. (2020). A tutorial on calibration measurements and calibration models for clinical prediction models. *Journal of the American Medical Informatics Association*, *27*(4), 621–633. https://doi.org/10.1093/jamia/ocz228

Huibers, M. J., Cohen, Z. D., Lemmens, L. H., Arntz, A., Peeters, F. P., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PLOS One*, *10*(11), e0140771. https://doi.org/10.1371/journal.pone.0140771

Jacobucci, R., Littlefield, A. K., Millner, A. J., Kleiman, E. M., & Steinley, D. (2020). Pairing machine learning and clinical psychology: How you evaluate predictive performance matters. https://doi.org/10.31234/osf.io/2yber

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. https://doi.org/10.7326/0003-4819-154-4-201102150-00006

Johansson, R., Sjoberg, E., Sjogren, M., Johnsson, E., Carlbring, P., Andersson, T., Rousseau, A., & Andersson, G. (2012). Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: Arandomized controlled trial. *PLOS One*, *7*(5), e36905. https://doi.org/10.1371/journal.pone.0036905

Joyner, M. J., & Paneth, N. (2019). Promises, promises, and precision medicine. *Journal of Clinical Investigation*, *129*(3), 946–948. https://doi.org/10.1172/JCI126119

Kapelner, A., Bleich, J., Levine, A., Cohen, Z. D., DeRubeis, R. J., & Berk, R. (2021). Evaluating the Effectiveness of Personalized Medicine with Software. *Frontiers in Big Data, 4*(8). doi: 10.3389/fdata.2021.572532

Karyotaki, E., Klein, A. M., Riper, H., Wit, L., Krijnen, L., Bol, E., Bolinski, F., Burger, S., Ebert, D. D., Auerbach, R. P., Kessler, R. C., Bruffaerts, R., Batelaan, N., van der Heijde, C. M., Vonk, P., Kleiboer, A., Wiers, R. W., & Cuijpers, P. (2019). Examining the effectiveness of a web-based intervention for symptoms of depression and anxiety in college students: Study protocol of a randomised controlled trial. *BMJ Open*, *9*(5), e028739. https://doi.org/10.1136/bmjopen-2018-028739

Keefe, J. R., Kim, T. T., DeRubeis, R. J., Streiner, D. L., Links, P. S., & McMain, S. F. (2020). Treatment selection in borderline personality disorder between dialectical behavior therapy and psychodynamic psychiatric management. *Psychological Medicine*, 1–9. https://doi.org/10.1017/S0033291720000550

Keefe, J. R., Wiltsey Stirman, S., Cohen, Z. D., DeRubeis, R. J., Smith, B. N., & Resick, P. A. (2018). In rape trauma PTSD, patient characteristics indicate which trauma-focused treatment they are most likely to complete. *Depression and anxiety*, *35*(4), 330–338. https://doi.org/10.1002/da.22731

Kent, D. M., Paulus, J. K., Van Klaveren, D., D'Agostino, R., Goodman, S., Hayward, R., Ioannidis, J. P., Patrick-Lake, B., Morton, S., & Pencina, M. (2020). The predictive approaches to treatment effect heterogeneity (PATH) statement. *Annals of Internal Medicine*, *172*(1), 35–45. https://doi.org/10.7326/M18-3667

Kent, D. M., Van Klaveren, D., Paulus, J. K., D'Agostino, R., Goodman, S., Hayward, R., Ioannidis, J. P., Patrick-Lake, B., Morton, S., & Pencina, M. (2020). The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement: Explanation and elaboration. *Annals of Internal Medicine*, *172*(1), W1–W25. https://doi.org/10.7326/M18-3668

Kerns, R. D., Burns, J. W., Shulman, M., Jensen, M. P., Nielson, W. R., Czlapinski, R., Dallas, M. I., Chatkoff, D., Sellinger, J., & Heapy, A. (2014). Can we improve cognitive–behavioral therapy for chronic back pain treatment engagement and adherence? A controlled trial of tailored versus standard therapy. *Health Psychology*, *33*(9), 938.

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

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & 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*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., Ebert, D. D., Hwang, I., Li, J., de Jonge, P., Nierenberg, A. A., Petukhova, M. V., Rosellini, A. J., Sampson, N. A., Schoevers, R. A., Wilcox, M. A., & Zaslavsky, A. M. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry*, *21*(10), 1366–1371. https://doi.org/10.1038/mp.2015.198

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

Kiekens, G., Hasking, P., Claes, L., Boyes, M., Mortier, P., Auerbach, R., Cuijpers, P., Demyttenaere, K., Green, J., & Kessler, R. (2019). Predicting the incidence of non-suicidal self-injury in college students. *European psychiatry*, *59*, 44–51. https://doi.org/10.1016/j.eurpsy.2019.04.002

Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin*, *65*(2), 110–136. https://doi.org/10.1037/h0022911

King, M., Walker, C., Levy, G., Bottomley, C., Royston, P., Weich, S., Bellon-Saameno, J. A., Moreno, B., Svab, I., Rotar, D., Rifel, J., Maaroos, H. I., Aluoja, A., Kalda, R., Neeleman, J., Geerlings, M. I., Xavier, M., Carraca, I., Goncalves-Pereira, M., Vicente, B., Saldivia, S., Melipillan, R., Torres-Gonzalez, F., & Nazareth, I. (2008). Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: The PredictD study. *Archives of General Psychiatry*, *65*(12), 1368–1376. https://doi.org/10.1001/archpsyc.65.12.1368

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

Kraepelien, M., Forsell, E., Karin, E., Johansson, R., Lindefors, N., & Kaldo, V. (2018). Comparing individually tailored to disorder-specific internet-based cognitive–behavioural therapy: Benchmarking study. *BJPsych Open*, *4*(4), 282–284. https://doi.org/10.1192/bjo.2018.41

Kraepelien, M., Svanborg, C., Lallerstedt, L., Sennerstam, V., Lindefors, N., & Kaldo, V. (2019). Individually tailored internet treatment in routine care: A feasibility study. *Internet Interventions*, *18*, 100263. https://doi.org/10.1016/j.invent.2019.100263

Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, *55*(4), 520 – 537. https://doi.org/10.1037/pst0000167

Lazarus, A. A. (2005). Multimodal therapy. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (2nd ed., pp. 105–120).

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

Levin, M. E., Haeger, J., & Cruz, R. A. (2018). Tailoring acceptance and commitment therapy skill coaching in the moment through smartphones: Results from a randomized controlled trial. *Mindfulness*, *10*(4), 689–699. https://doi.org/10.1007/s12671-018-1004-2

Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder.* Guilford Press.

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

Lorenzo-Luaces, L., DeRubeis, R. J., van Straten, A., & Tiemens, B. (2017). A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. *Journal of Affective Disorders*, *213*, 78–85. https://doi.org/10.1016/j.jad.2017.02.010

Lorenzo-Luaces, L., Peipert, A., De Jesús Romero, R., Rutter, L. A., & Rodriguez-Quintana, N. (2020). Personalized medicine and cognitive behavioral therapies for depression: Small effects, big problems, and bigger data. *International Journal of Cognitive Therapy*. https://doi.org/https://doi.org/10.1007/s41811-020-00094-3

Lorimer, B., Delgadillo, J., Kellett, S., & Lawrence, J. (2020). Dynamic prediction and identification of cases at risk of relapse following completion of low-intensity cognitive behavioural therapy. *Psychotherapy Research*, *31*(1), 1–14. https://doi.org/10.1080/10503307.2020.1733127

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

Luedtke, A. R., & van der Laan, M. J. (2016a). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of Statistics*, *44*(2), 713. https://doi.org/10.1214/15-AOS1384

Luedtke, A. R., & van der Laan, M. J. (2016b). Super-learning of an optimal dynamic treatment rule. *International Journal of Biostatistics*, *12*(1), 305–332. https://doi.org/10.1515/ijb-2015-0052

Luedtke, A. R., & van der Laan, M. J. (2017). Evaluating the impact of treating the optimal subgroup. *Statistical Methods in Medical Research*, *26*(4), 1630–1640. https://doi.org/10.1177/0962280217708664

Lutz, W., Rubel, J. A., Schwartz, B., Schilling, V., & Deisenhofer, A. K. (2019). Towards integrating personalized feedback research into clinical practice: Development of the

Trier Treatment Navigator (TTN). *Behaviour Research and Therapy*, *120*, 103438. https://doi.org/10.1016/j.brat.2019.103438

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

Lutz, W., Schwartz, B., Gomez Penedo, M. J., Boyle, K., & Deisenhofer, A. K. (2020). Working towards the development and implementation of precision mental healthcare: An example. *Administration and Policy in Mental Health*, *47*(5), 856–861. https://doi.org/10.1007/s10488-020-01053-y

Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-25953-0

Lutz, W., Deisenhofer, A.-K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2021). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology*, *in press*, https://doi.org/10.1037/ccp0000642.

Lynam, A. L., Dennis, J. M., Owen, K. R., Oram, R. A., Jones, A. G., Shields, B. M., & Ferrat, L. A. (2020). Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: Application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagnostic and Prognostic Research*, *4*, 6. https://doi.org/10.1186/s41512-020-00075-2

Macdonald, H., Loder, E., & Abbasi, K. (2020). Living systematic reviews at The BMJ. *British Medical Journal*, *370*, m2925. https://doi.org/10.1136/bmj.m2925

Maeda, Y., Ichikawa, R., Misawa, J., Shibuya, A., Hishiki, T., Maeda, T., Yoshino, A., & Kondo, Y. (2019). External validation of the TRISS, CRASH, and IMPACT prognostic models in severe traumatic brain injury in Japan. *PLOS One*, *14*(8), e0221791. https://doi.org/10.1371/journal.pone.0221791

Magnavita, J. J., & Lilienfeld, S. O. (2016). Clinical expertise and decision making: An overview of bias in clinical practice. In J. J. Magnavita (Ed.), *Clinical decision making in mental health practice* (pp. 23–60). American Psychological Association. https://doi.org/10.1037/14711-002

Malhi, G. S., Bassett, D., Boyce, P., Bryant, R., Fitzgerald, P. B., Fritz, K., Hopwood, M., Lyndon, B., Mulder, R., Murray, G., Porter, R., & Singh, A. B. (2015). Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Australian and New Zealand Journal of Psychiatry*, *49*(12), 1087–1206. https://doi.org/10.1177/0004867415617657

McCullough Jr, J. P. (2003). *Treatment for chronic depression: Cognitive behavioral analysis system of psychotherapy (CBASP)* (Vol. 13). Educational Publishing Foundation.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* University of Minnesota Press. https://doi.org/http://dx.doi.org/10.1037/11281-000

Miller, I. W., Keitner, G. I., Ryan, C. E., Solomon, D. A., Cardemil, E. V., & Beevers, C. G. (2005). Treatment matching in the posthospital care of depressed patients. *American Journal of Psychiatry*, *162*(11), 2131-2138. https://doi.org/10.1176/appi.ajp.162.11.2131

Mohr, D. C., Schueller, S. M., Tomasino, K. N., Kaiser, S. M., Alam, N., Karr, C., Vergara, J. L., Gray, E. L., Kwasny, M. J., & Lattie, E. G. (2019). Comparison of the Effects of Coaching and Receipt of App Recommendations on Depression, Anxiety, and Engagement in the IntelliCare Platform: Factorial Randomized Controlled Trial. *Journal of Medical Internet Research*, *21*(8), e13609. https://doi.org/10.2196/13609

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*, *162*(1), W1-W73. https://doi.org/10.7326/M14-0698

Moritz, S., Stepulovs, O., Schröder, J., Hottenrott, B., Meyer, B., & Hauschildt, M. (2016). Is the whole less than the sum of its parts? Full versus individually adapted metacognitive self-help for obsessive-compulsive disorder: A randomized controlled trial. *Journal of Obsessive-Compulsive and Related Disorders*, *9*, 107–115. https://doi.org/10.1016/j.jocrd.2016.04.001

Nahum-Shani, I., Hekler, E. B., & Spruijt-Metz, D. (2015). Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychology*, *34*(S), 1209-1219. https://doi.org/10.1037/hea0000306

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, *52*(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8

Najafabadi, A. H. Z., Ramspek, C. L., Dekker, F. W., Heus, P., Hooft, L., Moons, K. G., Peul, W. C., Collins, G. S., Steyerberg, E. W., & van Diepen, M. (2020). TRIPOD statement: A preliminary pre-post analysis of reporting and methods of prediction models. *BMJ Open*, *10*(9), e041537. https://doi.org/10.1136/bmjopen-2020-041537

National Collaborating Centre for Mental Health. (2009). *Borderline personality disorder: Treatment and management*. British Psychological Society.

National Collaborating Centre for Mental Health. (2011). *Common mental health disorders: Identification and pathways to care* (Vol. 123). RCPsych Publications.

Newton-Howes, G., Tyrer, P., Johnson, T., Mulder, R., Kool, S., Dekker, J., & Schoevers, R. (2014). Influence of personality on the outcome of treatment in depression: Systematic review and meta-analysis. *Journal of Personality Disorders*, *28*(4), 577–593. https://doi.org/10.1521/pedi_2013_27_070

Norcross, J. C. (2005). A primer on psychotherapy integration. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of Psychotherapy Integration (*2nd ed., pp. 3–23).

Norcross, J. C., & Prochaska, J. O. (1988). A study of eclectic (and integrative) views revisited. *Professional Psychology: Research and Practice*, *19*(2), 170–174. https://doi.org/10.1037/0735-7028.19.2.170

Nordgren, L. B., Hedman, E., Etienne, J., Bodin, J., Kadowaki, A., Eriksson, S., Lindkvist, E., Andersson, G., & Carlbring, P. (2014). Effectiveness and cost-effectiveness of individually tailored internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: A randomized controlled trial. *Behaviour Research and Therapy*, *59*, 1–11. https://doi.org/10.1016/j.brat.2014.05.007

Nusinovici, S., Tham, Y. C., Chak Yan, M. Y., Wei Ting, D. S., Li, J., Sabanayagam, C., Wong, T. Y., & Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of Clinical Epidemiology*, *122*, 56–69. https://doi.org/10.1016/j.jclinepi.2020.03.002

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447. https://doi.org/10.1126/science.aax2342

Ogrodniczuk, J. S., Joyce, A. S., & Piper, W. E. (2005). Strategies for reducing patient-initiated premature termination of psychotherapy. *Harvard Review of Psychiatry*, *13*(2), 57–70. https://doi.org/10.1080/10673220590956429

Ong, W. T., Murphy, D., & Joseph, S. (2020). Unnecessary and incompatible: A critical response to Cooper and McLeod's conceptualization of a pluralistic framework for person-centered therapy. *Person-Centered & Experiential Psychotherapies*, *19*(2), 168–182. https://doi.org/10.1080/14779757.2020.1717987

Papakostas, G. I., & Fava, M. (2008). Predictors, moderators, and mediators (correlates) of treatment outcome in major depressive disorder. *Dialogues in Clinical Neuroscience*, *10*(4), 439–451. https://doi.org/10.31887/DCNS.2008.10.4/gipapakostas

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*(2), 109–118. https://doi.org/10.1037/h0024436

Pearson, R., Pisner, D., Meyer, B., Shumake, J., & Beevers, C. G. (2019). A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression. *Psychological Medicine*, *49*(14), 2330–2341. https://doi.org/10.1017/S003329171800315X

Pencina, M. J., Goldstein, B. A., & D'Agostino, R. B. (2020). Prediction models-development, evaluation, and clinical application. *New England Journal of Medicine*, *382*(17), 1583–1586. https://doi.org/10.1056/NEJMp2000589

Perlis, R. H. (2016). Abandoning personalization to get to precision in the pharmacotherapy of depression. *World Psychiatry*, *15*(3), 228–235. https://doi.org/10.1002/wps.20345

Persons, J. B. (1991). Psychotherapy outcome studies do not accurately represent current models of psychotherapy: A proposed remedy. *American psychologist*, *46*(2), 99–106. https://doi.org/10.1037/0003-066X.46.2.99

Petkova, E., Park, H., Ciarleglio, A., Todd Ogden, R., & Tarpey, T. (2019). Optimising treatment decision rules through generated effect modifiers: A precision medicine tutorial. *BJPsych Open*, *6*(1), e2. https://doi.org/10.1192/bjo.2019.85

Petkova, E., Tarpey, T., Su, Z., & Ogden, R. T. (2017). Generated effect modifiers (GEM's) in randomized clinical trials. *Biostatistics*, *18*(1), 105–118. https://doi.org/10.1093/biostatistics/kxw035

Phung, M. T., Tin, S. T., & Elwood, J. M. (2019). Prognostic models for breast cancer: A systematic review. *BMC Cancer*, *19*(1), 230. https://doi.org/10.1186/s12885-019-5442-6

Piet, J., & Hougaard, E. (2011). The effect of mindfulness-based cognitive therapy for prevention of relapse in recurrent major depressive disorder: A systematic review and meta-analysis. *Clinical Psychology Review*, *31*(6), 1032–1040. https://doi.org/10.1016/j.cpr.2011.05.002

Prasad, V., Fojo, T., & Brada, M. (2016). Precision oncology: Origins, optimism, and potential. *Lancet Oncology*, *17*(2), e81–e86. https://doi.org/10.1016/S1470-2045(15)00620-8

Project MATCH Research Group. (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*, *58*(1), 7–29. https://www.ncbi.nlm.nih.gov/pubmed/8979210

Project MATCH Research Group. (1998). Matching patients with alcohol disorders to treatments: Clinical implications from Project MATCH. *Journal of Mental Health*, *7*(6), 589–602. https://doi.org/10.1080/09638239817743

Rekkas, A., Paulus, J. K., Raman, G., Wong, J. B., Steyerberg, E. W., Rijnbeek, P. R., Kent, D. M., & van Klaveren, D. (2019). Predictive approaches to heterogeneous treatment effects: A systematic review. *medRxiv*. https://doi.org/10.1101/19010827

Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Jr., Martin, G. P., Reitsma, J. B., Moons, K. G. M., Collins, G., & van Smeden, M. (2020). Calculating the sample size required for developing a clinical prediction model. *British Medical Journal*, *368*, m441. https://doi.org/10.1136/bmj.m441

Riley, R. D., Hayden, J. A., Steyerberg, E. W., Moons, K. G., Abrams, K., Kyzas, P. A., Malats, N., Briggs, A., Schroter, S., Altman, D. G., Hemingway, H., & Group, P. (2013). Prognosis Research Strategy (PROGRESS) 2: Prognostic factor research. *PLOS Medicine*, *10*(2), e1001380. https://doi.org/10.1371/journal.pmed.1001380

Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins, G. S. (2019a). Minimum sample size for developing a multivariable prediction model: Part I– Continuous outcomes. *Statistics in Medicine*, *38*(7), 1262–1275.

Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins, G. S. (2019b). Minimum sample size for developing a multivariable prediction model: PART II-binary and time-to-event outcomes. *Statistics in Medicine*, *38*(7), 1276–1296. https://doi.org/10.1002/sim.7993

Riley, R. D., van der Windt, D., Croft, P., & Moons, K. G. (2019). *Prognosis Research in Healthcare: Concepts, Methods, and Impact*. Oxford University Press.

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2009). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575. https://doi.org/10.1109/TPAMI.2009.187.

Rosellini, A. J., Dussaillant, F., Zubizarreta, J. R., Kessler, R. C., & Rose, S. (2018). Predicting posttraumatic stress disorder following a natural disaster. *Journal of Psychiatric Research*, *96*, 15–22. https://doi.org/10.1016/j.jpsychires.2017.09.010

Rubel, J. A., Fisher, A. J., Husen, K., & Lutz, W. (2018). Translating person-specific network models into personalized treatments: Development and demonstration of the Dynamic Assessment Treatment Algorithm for Individual Networks (DATA-IN). *Psychotherapy and Psychosomatics*, *87*(4), 249–251. https://doi.org/10.1159/000487769

Rubel, J. A., Zilcha-Mano, S., Giesemann, J., Prinz, J., & Lutz, W. (2020). Predicting personalized process-outcome associations in psychotherapy using machine learning approaches-A demonstration. *Psychotherapy Research*, *30*(3), 300–309. https://doi.org/10.1080/10503307.2019.1597994

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Stewart, J. W., Nierenberg, A. A., Thase, M. E., Ritz, L., Biggs, M. M., Warden, D., Luther, J. F., Shores-Wilson, K., Niederehe, G., Fava, M., & Team, S. D. S. (2006). Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *New England Journal of Medicine*, *354*(12), 1231–1242. https://doi.org/10.1056/NEJMoa052963

Ryle, A., Poynton, A. M., & Brockman, B. J. (1990). *Cognitive-analytic therapy: Active participation in change: A new integration in brief psychotherapy*. John Wiley & Sons.

Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., Baldwin, H., Danese, A., Fazel, S., Steyerberg, E. W., Stahl, D., & Fusar-Poli, P. (2020). Implementing precision psychiatry: A systematic review of individualized prediction models for clinical practice. *Schizophrenia Bulletin,* sbaa120. https://doi.org/10.1093/schbul/sbaa120

Sauerbrei, W., Perperoglou, A., Schmid, M., Abrahamowicz, M., Becher, H., Binder, H., Dunkler, D., Harrell, F. E., Jr., Royston, P., Heinze, G., & For TG2 of the STRATOS Initiative (2020). State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. *Diagnostic and Prognostic Research*, *4*, 3. https://doi.org/10.1186/s41512-020-00074-3

Saunders, R., Buckman, J. E. J., & Pilling, S. (2020). Latent variable mixture modelling and individual treatment prediction. *Behaviour Research and Therapy*, *124*, 103505. https://doi.org/10.1016/j.brat.2019.103505

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. https://doi.org/10.1016/j.jad.2016.03.011

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin*, *66*(3), 178–200. https://doi.org/10.1037/h0023624

Schon, D. A. (1984). *The reflective practitioner: How professionals think in action* (Vol. 5126). Basic Books.

Schulte, D., Künzel, R., Pepping, G., & Schulte-Bahrenberg, T. (1992). Tailor-made versus standardized therapy of phobic patients. *Advances in Behaviour Research and Therapy*, *14*(2), 67–92. https://doi.org/10.1016/0146-6402(92)90001-5

Schwartz, B., Cohen, Z. D., Rubel, J. A., Zimmermann, D., Wittmann, W. W., & Lutz, W. (2020). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research*, *31*(1), 1–19. https://doi.org/10.1080/10503307.2020.1769219

Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2001). *Mindfulness-based Cognitive Therapy for Depression: A new approach to preventing relapse*. Guilford Publications.

Senn, S. (2018). Statistical pitfalls of personalized medicine. *Nature*, *563*(7733), 619–621. https://doi.org/10.1038/d41586-018-07535-2

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330

Silfvernagel, K., Carlbring, P., Kabo, J., Edstrom, S., Eriksson, J., Manson, L., & Andersson, G. (2012). Individually tailored internet-based treatment for young adults and adults with panic attacks: Randomized controlled trial. *Journal of Medical Internet Research*, *14*(3), e65. https://doi.org/10.2196/jmir.1853

Silfvernagel, K., Westlinder, A., Andersson, S., Bergman, K., Diaz Hernandez, R., Fallhagen, L., Lundqvist, I., Masri, N., Viberg, L., Forsberg, M. L., Lind, M., Berger, T., Carlbring, P., & Andersson, G. (2018). Individually tailored internet-based cognitive behaviour therapy for older adults with anxiety and depression: A randomised controlled trial. *Cognitive Behaviour Therapy*, *47*(4), 286–300. https://doi.org/10.1080/16506073.2017.1388276

Simon, G. E., & Perlis, R. H. (2010). Personalized medicine for depression: Can we match patients with treatments? *American Journal of Psychiatry*, *167*(12), 1445–1455. https://doi.org/10.1176/appi.ajp.2010.09111680

Siontis, G. C., Tzoulaki, I., Castaldi, P. J., & Ioannidis, J. P. (2015). External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *Journal of Clinical Epidemiology*, *68*(1), 25–34. https://doi.org/10.1016/j.jclinepi.2014.09.007

Smith, B., & Sechrest, L. (1991). Treatment of aptitude X treatment interactions. *Journal of Consulting and Clinical Psychology*, *59*(2), 233–244. https://doi.org/10.1037//0022-006x.59.2.233

Smits, J. A., Powers, M. B., & Otto, M. W. (2019). *Personalized Exposure Therapy: A person-centered transdiagnostic approach*. Oxford University Press.

Smyth, J. M., & Heron, K. E. (2016). Is providing mobile interventions "just-in-time" helpful? An experimental proof of concept study of just-in-time intervention for stress management. *2016 IEEE Wireless Health*, 1–7. https://doi.org/doi: 10.1109/WH.2016.7764561

Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, *59*(2), 205–216. https://doi.org/10.1037/0022-006X.59.2.205

Soto, A., Smith, T. B., Griner, D., Domenech Rodriguez, M., & Bernal, G. (2018). Cultural adaptations and therapist multicultural competence: Two meta-analytic reviews. *Journal of Clinical Psychology*, *74*(11), 1907–1923. https://doi.org/10.1002/jclp.22679

Souery, D., Serretti, A., Calati, R., Oswald, P., Massat, I., Konstantinidis, A., Linotte, S., Bollen, J., Demyttenaere, K., Kasper, S., Lecrubier, Y., Montgomery, S., Zohar, J., & Mendlewicz, J. (2011). Switching antidepressant class does not improve response or remission in treatment-resistant depression. *Journal of Clinical Psychopharmacology*, *31*(4), 512–516. https://doi.org/10.1097/JCP.0b013e3182228619

Stacey, D., Legare, F., Lewis, K., Barry, M. J., Bennett, C. L., Eden, K. B., Holmes-Rovner, M., Llewellyn-Thomas, H., Lyddiatt, A., Thomson, R., & Trevena, L. (2017). Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews*, *4*(4), CD001431. https://doi.org/10.1002/14651858.CD001431.pub5

Steyerberg, E. W. (2019). *Clinical Prediction Models*. Springer.

Steyerberg, E. W., Bleeker, S. E., Moll, H. A., Grobbee, D. E., & Moons, K. G. M. (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*, *56*(5), 441–447. https://doi.org/10.1016/s0895-4356(03)00047-7

Steyerberg, E. W., & Harrell, F. E., Jr. (2016). Prediction models need appropriate internal, internal-external, and external validation. *Journal of Clinical Epidemiology*, *69*, 245–247. https://doi.org/10.1016/j.jclinepi.2015.04.005

Steyerberg, E. W., Harrell, F. E., Borsboom, G. J. J. M., Eijkemans, M. J. C., Vergouwe, Y., & Habbema, J. D. F. (2001). Internal validation of predictive models. *Journal of Clinical Epidemiology*, *54*(8), 774–781. https://doi.org/10.1016/s0895-4356(01)00341-9

Steyerberg, E. W., Moons, K. G., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., Riley, R. D., Hemingway, H., Altman, D. G., & Group, P. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. *PLOS Medicine*, *10*(2), e1001381. https://doi.org/10.1371/journal.pmed.1001381

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*, *21*(1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Strandskov, S. W., Ghaderi, A., Andersson, H., Parmskog, N., Hjort, E., Warn, A. S., Jannert, M., & Andersson, G. (2017). Effects of tailored and ACT-influenced Internet-based CBT for Eating Disorders and the relation between knowledge acquisition and outcome: A Randomized Controlled Trial. *Behavior Therapy*, *48*(5), 624–637. https://doi.org/10.1016/j.beth.2017.02.002

Strupp, H. H. (1964). A bibliography of research in psychotherapy. *Psychotherapy Research Project*, Department of Psychiatry, University of North Carolina School of Medicine.

Symons, M., Feeney, G. F. X., Gallagher, M. R., Young, R. M., & Connor, J. P. (2020). Predicting alcohol dependence treatment outcomes: A prospective comparative study of clinical psychologists versus 'trained' machine learning models. *Addiction*, *115*(11), 2164–2175. https://doi.org/10.1111/add.15038

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. *Clinical Psychology: Science and Practice*, *22*(4), 317–338. https://doi.org/10.1111/cpsp.12122

Trivedi, M. H., Rush, A. J., Crismon, M. L., Kashner, T. M., Toprac, M. G., Carmody, T. J., Key, T., Biggs, M. M., Shores-Wilson, K., Witte, B., Suppes, T., Miller, A. L., Altshuler, K. Z., & Shon, S. P. (2004). Clinical results for patients with major depressive disorder in the Texas Medication Algorithm Project. *Archives of General Psychiatry*, *61*(7), 669–680. https://doi.org/10.1001/archpsyc.61.7.669

Twomey, C., O'Reilly, G., & Meyer, B. (2017). Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: A meta-analysis. *Psychiatry Research*, *256*, 371–377. https://doi.org/10.1016/j.psychres.2017.06.081

Uwatoko, T., Luo, Y., Sakata, M., Kobayashi, D., Sakagami, Y., Takemoto, K., Collins, L. M., Watkins, E., Hollon, S. D., Wason, J., Noma, H., Horikoshi, M., Kawamura, T., Iwami, T., & Furukawa, T. A. (2018). Healthy Campus Trial: A multiphase optimization strategy (MOST) fully factorial trial to optimize the smartphone cognitive behavioral therapy (CBT) app for mental health promotion among university students: Study protocol for a randomized controlled trial. *Trials*, *19*(1), 353. https://doi.org/10.1186/s13063-018-2719-z

van Beugen, S., Ferwerda, M., Spillekom-van Koulil, S., Smit, J. V., Zeeuwen-Franssen, M. E., Kroft, E. B., de Jong, E. M., Otero, M. E., Donders, A. R., van de Kerkhof, P. C., van Middendorp, H., & Evers, A. W. (2016). Tailored therapist-guided internet-based cognitive behavioral treatment for psoriasis: A randomized controlled trial. *Psychotherapy Psychosomatics*, *85*(5), 297–307. https://doi.org/10.1159/000447267

van Bronswijk, S. C., Bruijniks, S. J., Lorenzo-Luaces, L., Derubeis, R. J., Lemmens, L. H., Peeters, F. P., & Huibers, M. J. (2020). Cross-trial prediction in psychotherapy: External validation of the Personalized Advantage Index using machine learning in two Dutch randomized trials comparing CBT versus IPT for depression. *Psychotherapy Research*, *31*(1), 1–14. https://doi.org/10.1080/10503307.2020.1823029

van Bronswijk, S. C., DeRubeis, R. J., Lemmens, L., Peeters, F., Keefe, J. R., Cohen, Z. D., & Huibers, M. J. H. (2019). Precision medicine for long-term depression outcomes using the Personalized Advantage Index approach: Cognitive therapy or interpersonal psychotherapy? *Psychological Medicine*, 1–11. https://doi.org/10.1017/S0033291719003192

van Bronswijk, S. C., Lemmens, L. H., Huibers, M. J., & Peeters, F. P. (2020). Selecting the optimal treatment for a depressed individual: Clinical judgment or statistical prediction? *Journal of Affective Disorders, 279*, 149–157. https://doi.org/10.1016/j.jad.2020.09.135

van Calster, B., Wynants, L., Timmerman, D., Steyerberg, E. W., & Collins, G. S. (2019). Predictive analytics in health care: How can we know it works? *Journal of the American Medical Informatics Association*, *26*(12), 1651–1654. https://doi.org/10.1093/jamia/ocz130

van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology*, *6*(1), 25. https://doi.org/10.2202/1544–6115.1309

van Smeden, M., Moons, K. G., de Groot, J. A., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research*, *28*(8), 2455–2474. https://doi.org/10.1177/0962280218784726

Varadhan, R., Segal, J. B., Boyd, C. M., Wu, A. W., & Weiss, C. O. (2013). A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *Journal of Clinical Epidemiology*, *66*(8), 818–825. https://doi.org/10.1016/j.jclinepi.2013.02.009

Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, *180*(Pt A), 68–77. https://doi.org/10.1016/j.neuroimage.2017.06.061

Vittengl, J. R., Anna Clark, L., 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. https://doi.org/10.1016/j.psychres.2017.03.032

Vollmer, S., Mateen, B. A., Bohner, G., Kiraly, F. J., Ghani, R., Jonsson, P., Cumbers, S., Jonas, A., McAllister, K. S. L., Myles, P., Granger, D., Birse, M., Branson, R., Moons, K. G. M., Collins, G. S., Ioannidis, J. P. A., Holmes, C., & Hemingway, H. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *British Medical Journal*, *368*, l6927. https://doi.org/10.1136/bmj.l6927

Walitzer, K. S., Dermen, K. H., & Connors, G. J. (1999). Strategies for preparing clients for treatment: A review. *Behavior Modification*, *23*(1), 129–151. https://doi.org/10.1177/0145445599231006

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

Walter, S., Turner, R., Macaskill, P., McCaffery, K., & Irwig, L. (2017). Beyond the treatment effect: Evaluating the effects of patient preferences in randomised trials. *Statistical Methods in Medical Research*, *26*(1), 489–507. https://doi.org/10.1177/0962280214550516

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. Routledge.

Wardenaar, K. J., van Loo, H. M., Cai, T., Fava, M., Gruber, M. J., Li, J., de Jonge, P., Nierenberg, A. A., Petukhova, M. V., Rose, S., Sampson, N. A., Schoevers, R. A., Wilcox, M. A., Alonso, J., Bromet, E. J., Bunting, B., Florescu, S. E., Fukao, A., Gureje, O., … Kessler, R. C. (2014). The effects of co-morbidity in defining major depression subtypes associated with long-term course and severity. *Psychological Medicine*, *44*(15), 3289–3302. https://doi.org/10.1017/S0033291714000993

Watzke, B., Ruddel, H., Jurgensen, R., Koch, U., Kriston, L., Grothgar, B., & Schulz, H. (2010). Effectiveness of systematic treatment selection for psychodynamic and cognitive-behavioural therapy: Randomised controlled trial in routine mental healthcare. *British Journal of Psychiatry*, *197*(2), 96-105. https://doi.org/10.1192/bjp.bp.109.072835

Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2019). Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology*, *88*(1), 25. https://doi.org/10.1037/ccp0000451

Webb, C. A., Trivedi, M. H., Cohen, Z. D., Dillon, D. G., Fournier, J. C., Goer, F., Fava, M., McGrath, P. J., Weissman, M., Parsey, R., Adams, P., Trombello, J. M., Cooper, C., Deldin, P., Oquendo, M. A., McInnis, M. G., Huys, Q., Bruder, G., Kurian, B. T., Jha, M., DeRubeis, R. J., & Pizzagalli, D. A. (2019). Personalized prediction of antidepressant v. placebo response: Evidence from the EMBARC study. *Psychological Medicine*, *49*(7), 1118–1127. https://doi.org/10.1017/S0033291718001708

Weisel, K. K., Zarski, A. C., Berger, T., Krieger, T., Schaub, M. P., Moser, C. T., Berking, M., Dey, M., Botella, C., Banos, R., Herrero, R., Etchemendy, E., Riper, H., Cuijpers, P., Bolinski, F., Kleiboer, A., Gorlich, D., Beecham, J., Jacobi, C., & Ebert, D. D. (2019). Efficacy and cost-effectiveness of guided and unguided internet- and mobile-based indicated transdiagnostic prevention of depression and anxiety (ICare Prevent): A three-armed randomized controlled trial in four European countries. *Internet Interventions*, *16*, 52–64. https://doi.org/10.1016/j.invent.2018.04.002

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., Daleiden, E. L., Ugueto, A. M., Ho, A., Martin, J., Gray, J., Alleyne, A., Langer, D. A., Southam-Gerow, M. A., Gibbons, R. D., & the Research Network on Youth Mental Health (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: a randomized effectiveness trial. *Archives of General Psychiatry*, *69*(3), 274–282. https://doi.org/10.1001/archgenpsychiatry.2011.147

Welsh, I., & Lyons, C. M. (2001). Evidence-based care and the case for intuition and tacit knowledge in clinical assessment and decision making in mental health nursing practice: An empirical contribution to the debate. *Journal of Psychiatric and Mental Health Nursing*, *8*(4), 299–305. https://doi.org/10.1046/j.1365-2850.2001.00386.x

Wessler, B. S., Lai YH, L., Kramer, W., Cangelosi, M., Raman, G., Lutz, J. S., & Kent, D. M. (2015). Clinical prediction models for cardiovascular disease: Tufts predictive analytics and comparative effectiveness clinical prediction model database. *Circulation: Cardiovascular Quality and Outcomes*, *8*(4), 368–375. https://doi.org/10.1161/CIRCOUTCOMES.115.001693

Wheeler, S. (1993). 17 Reservations about eclectic and integrative approaches to counselling. *Questions and Answers on Counselling in Action*, *16*, 86.

Wilkinson, J., Arnold, K. F., Murray, E. J., van Smeden, M., Carr, K., Sippy, R., de Kamps, M., Beam, A., Konigorski, S., & Lippert, C. (2020). Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health, 2*(12), e677–e680. https://doi.org/10.1016/S2589-7500(20)30200-4

Windle, E., Tee, H., Sabitova, A., Jovanovic, N., Priebe, S., & Carr, C. (2019). Association of patient treatment preference with dropout and clinical outcomes in adult psychosocial

mental health interventions: A systematic review and meta-analysis. *JAMA Psychiatry, 77*(3), 294-302. https://doi.org/10.1001/jamapsychiatry.2019.3750

Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, *16*, 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., Bonten, M. M. J., Damen, J. A. A., Debray, T. P. A., De Vos, M., Dhiman, P., Haller, M. C., Harhay, M. O., Henckaerts, L., Kreuzberger, N., Lohman, A., Luijken, K., Ma, J., Andaur, C. L., … van Smeden, M. (2020). Prediction models for diagnosis and prognosis of Covid-19 infection: Systematic review and critical appraisal. *British Medical Journal*, *369*, m1328. https://doi.org/10.1136/bmj.m1328

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*(12), e1584-e1590. https://doi.org/10.4088/JCP.15m10081

Zoellner, L. A., Roy-Byrne, P. P., Mavissakalian, M., & Feeny, N. C. (2019). Doubly randomized preference trial of prolonged exposure versus sertraline for treatment of PTSD. *American Journal of Psychiatry*, *176*(4), 287–296. https://doi.org/10.1176/appi.ajp.2018.17090995