Behaviour Research and Therapy 172 (2024) 104443

Contents lists available at ScienceDirect

Behaviour Research and Therapy

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

# Implementing precision methods in personalizing psychological therapies: Barriers and possible ways forward

Anne-Katharina Deisenhofer [a,*], Michael Barkham [b], Esther T. Beierl [c], Brian Schwartz [a], Katie Aafjes-van Doorn [d], Christopher G. Beevers [e], Isabel M. Berwian [f], Simon E. Blackwell [g], Claudi L. Bockting [h], Eva-Lotta Brakemeier [i], Gary Brown [j], Joshua E.J. Buckman [k], Louis G. Castonguay [l], Claire E. Cusack [m], Tim Dalgleish [n], Kim de Jong [o], Jaime Delgadillo [b], Robert J. DeRubeis [p], Ellen Driessen [q], Jill Ehrenreich-May [r], Aaron J. Fisher [s], Eiko I. Fried [t], Jessica Fritz [n,u], Toshi A. Furukawa [v], Claire M. Gillan [w], Juan Martín Gómez Penedo [x], Peter F. Hitchcock [y], Stefan G. Hofmann [u], Steven D. Hollon [z], Nicholas C. Jacobson [aa], Daniel R. Karlin [ab], Chi Tak Lee [ac], Cheri A. Levinson [m], Lorenzo Lorenzo-Luaces [ad], Riley McDanal [ae], Danilo Moggia [a], Mei Yi Ng [af], Lesley A. Norris [ag], Vikram Patel [ah], Marilyn L. Piccirillo [ai], Stephen Pilling [k], Julian A. Rubel [aj], Gonzalo Salazar-de-Pablo [ak], Rob Saunders [k], Jessica L. Schleider [al], Paula P. Schnurr [am], Stephen M. Schueller [an], Greg J. Siegle [ao], Rudolf Uher [ap], Ed Watkins [aq], Christian A. Webb [ar], Shannon Wiltsey Stirman [as], Laure Wynants [at], Soo Jeong Youn [au], Sigal Zilcha-Mano [av], Wolfgang Lutz [a], Zachary D. Cohen [aw,**]

[a] University of Trier, Germany
[b] University of Sheffield, UK
[c] University of Oxford, UK
[d] Yeshiva University, USA
[e] The University of Texas at Austin, USA
[f] Princeton University, USA
[g] Ruhr University Bochum, Germany
[h] AmsterdamUMC, Department of Psychiatry, Research Program Amsterdam Public Health and Centre for Urban Mental Health, University of Amsterdam, the Netherlands
[i] University of Greifswald, Germany
[j] Royal Holloway University of London, UK
[k] University College London, UK
[l] Pennsylvania State University, USA
[m] University of Louisville, USA
[n] University of Cambridge, UK
[o] Leiden University, Institute of Psychology, the Netherlands
[p] University of Pennsylvania, USA
[q] Radboud University, the Netherlands
[r] University of Miami, USA
[s] University of California, Berkeley, USA
[t] Leiden University, the Netherlands
[u] Philipps University of Marburg, Germany
[v] Kyoto University, Japan
[w] School of Psychology, Trinity College Institute for Neuroscience, And Global Brain Health Institute, Trinity College Dublin, Ireland
[x] University of Buenos Aires, Argentina
[y] Emory University, USA
[z] Vanderbilt University, USA
[aa] Dartmouth Geisel School of Medicine, USA

* Corresponding author. Clinical Psychology and Psychotherapy, Department of Psychology, University of Trier, D-54296, Trier, Germany.
** Corresponding author. Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, USA.
*E-mail addresses:* deisenhofer@uni-trier.de (A.-K. Deisenhofer), cohenzd@arizona.edu (Z.D. Cohen).

https://doi.org/10.1016/j.brat.2023.104443
Received 23 August 2023; Received in revised form 21 November 2023; Accepted 27 November 2023
Available online 1 December 2023
0005-7967/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[ab] MindMed, Inc., Tufts University School of Medicine, USA
[ac] Trinity College Dublin, Ireland
[ad] Indiana University-Bloomington, USA
[ae] Stony Brook University, USA
[af] Florida International University, USA
[ag] Warren Alpert Medical School of Brown University, USA
[ah] Harvard Medical School, USA
[ai] University of Washington, USA
[aj] Osnabrück University, Germany
[ak] King's College London and South London and Maudsley NHS Foundation Trust, UK
[al] Stony Brook University and Feinberg School of Medicine Northwestern University, USA
[am] National Center for PTSD and Geisel School of Medicine at Dartmouth, USA
[an] University of California, Irvine, USA
[ao] University of Pittsburgh, USA
[ap] Dalhousie University, Canada
[aq] University of Exeter, UK
[ar] Harvard Medical School, McLean Hospital, USA
[as] National Center for PTSD and Stanford University, USA
[at] Maastricht University, the Netherlands
[au] Reliant Medical Group, OptumCare and Harvard Medical School, USA
[av] University of Haifa, Israel
[aw] University of California, Los Angeles and University of Arizona, USA

ARTICLE INFO

*Keywords:*
Precision mental health care methods
Personalizing psychological treatments
Data-driven psychological therapies
Implementation barriers

ABSTRACT

Personalization of psychological therapies has always been used by clinicians and describes all efforts to select, adjust, or modify a treatment for the individual to improve outcomes. Precision mental health care approaches can be considered under the umbrella term personalization and specify methods that are algorithmic, quantitative, and empirically derived. Despite a growing research literature demonstrating the efficacy of these approaches, they are rarely tested in clinical practice. A statistically optimized, targeted clinical recommendation is not by itself sufficient to influence clinical practice in a beneficial way; barriers related to dissemination and implementation require increased attention. This article describes clinical and practical factors, technical aspects, statistical considerations, and fundamental contextual issues that should be considered to facilitate data-driven treatments in mental health care contexts in clinical practice.

Precision mental health care and other treatment personalization approaches are promising pathways for targeting interventions to specific individuals based on their personal characteristics, thereby having the potential for enhancing client outcomes (Cohen & DeRubeis, 2018). Although evidence from meta-analytic (Nye, Delgadillo, & Barkham, 2023) and empirical studies (e.g., Delgadillo et al., 2022) indicate that results yield gains that can, at best, be described in terms of small effects, such benefits when delivered at scale can have considerable impact at a population level (Barkham, 2023). It is in such a context (i.e., allocating people to the best-fitting treatment), that the current article builds on Gordon Paul's litany "*What* treatment, by *whom*, is most effective for *this* individual with *that* specific problem, and under *which* set of circumstances?" (Paul, 1967, p. 111) by focusing on issues of implementation.

In the context of treatment, the use of the terms 'precision' and 'personalization' has evolved over time (Tran, Klossner, Crain, & Prasad, 2020). The result has been conflicting interpretations leading to ongoing debate, with definitions of 'precision medicine' and 'personalized medicine' often being indistinguishable (e.g., Szatmari & Susser, 2022). However, with the emphasis on the quantitative foundations of this approach, the field has shifted away from "personalization" (Perlis, 2016) toward the term "precision medicine" (Collins & Varmus, 2015). Although it is likely that the terms are used by some people interchangeably or without any strongly held preference for one term over another, we take the view, first, that the terms *can* be differentiated, and second, that by doing so, clearer accounts of this developing field can be generated, especially in relation to psychological treatments.

Our starting point in the current article is that the term *precision* indicates the involvement of data and that its use should be reserved for approaches that are algorithmic, quantitative, and empirically derived. The application of precision methods to the field of clinical psychology and psychiatry has led to *data-informed and data-driven psychological therapies*, in which decision rules for the treatment of the client can be created and implemented at the beginning (i.e., baseline; Cohen, Delgadillo, & DeRubeis, 2021). At the same time, it may also mean that in an ongoing psychological treatment, predictions are made based on dynamic prediction systems (i.e., where the predictions are constantly updated by the collection of new data using repeated measurements; sometimes called just-in-time adaptive interventions). It may also mean that a digital therapy is automatically adapted to a patient's needs based on data.

By contrast, the term *personalized* can be considered an umbrella term that describes any effort by the clinician to select, adapt, or adjust treatment to an individual with the goal of improving outcomes, including approaches that originate from and are driven by a holistic and person-centered approach to an individual including, for example, client preferences (e.g., Norcross & Cooper, 2021). It is also consistent with the well-established clinical practice of case-formulation, as described by Persons (2008) who has also incorporated the use of progress monitoring (Persons, Koerner, Eidelman, Thomas, & Liu, 2016). In addition, personalized approaches also include treatment guidelines for matching patients to disorder-specific protocols (see section titled "Matching Treatments to Disorders" in Cohen et al., 2021, pp. 678–679 for further discussion) and newer approaches that focus on selecting specific evidence-based practices for particular biopsychosocial processes. An example of the latter approach is process-based therapy (Hayes et al., 2019; Hofmann & Hayes, 2019). In addition to personalization, other terms are also used in the literature to describe changes to the treatment protocol. The interested reader is referred to the review by Bennett and Shafran (2023). As this article focuses on data-driven (and data-informed) approaches, no further differentiation will be made here. Personalization is thus not a new approach but has always been practiced by clinicians to tailor psychological therapies to individuals. With technological advances, statistical innovations, and scientific developments, precision mental health care is

advancing and finding its way into practice in order to inform treatment personalization (Lutz, Deisenhofer, et al., 2022; Schwartz et al., 2021).

So far, approaches based on statistical modeling have been viewed with considerable promise for psychological therapies and have been applied across a variety of contexts and disorders (Chekroud et al., 2021). One key approach has been machine learning, which refers to a class of approaches that connects computer science and statistics, and a field that has developed considerably over recent years. The primary objective of machine learning is to enable computers to make successful automated predictions while learning from past experiences and thus transforming data into actionable knowledge. Cohen et al. (2021) outlined different ways in which precision mental health care approaches can inform treatment including, but not limited to, identifying the optimal level of care (Delgadillo et al., 2022), selecting a particular treatment for patients in a personalized way (Cohen et al., 2022; Moggia, Saxon, Lutz, Hardy, & Barkham, 2023), selecting or ordering particular intervention strategies within a particular treatment for a given client (Fisher et al., 2019; Pearson, Pisner, Meyer, Shumake, & Beevers, 2019; Rubel, Fisher, Husen, & Lutz, 2018), and selecting clinicians for specific clients (Constantino, Boswell, Coyne, Swales, & Kraus, 2021; Delgadillo, Rubel, & Barkham, 2020).

The evidence supporting most approaches has been limited, hampered by problems such as insufficient sample size (Lorenzo-Luaces, Peipert, De Jesús Romero, Rutter, & Rodriguez-Quintana, 2021; Luedtke, Sadikova, & Kessler, 2019) and lack of proper external validation (Meehan et al., 2022; Salazar de Pablo et al., 2020). As a result, implementation in clinical practice has been limited, resulting in a gap between the scientific developments in precision methods and the clinical care received by clients. The existence of a gap is not unique to this specific field, but rather typical of the trajectory from research production to adoption in practice (Greenhalgh, 2008; Robert, Greenhalgh, MacFarlane, & Peacock, 2010). Accordingly, implementation science has the potential to enable us to better understand the barriers and facilitators to future uptake of precision methods in the clinical practice of psychological treatments. Relevant bodies of literature pertain to proposed translational frameworks (e.g., Chambers, Feero, & Khoury, 2016; Chanfreau-Coffinier et al., 2019; Ward & Ginsburg, 2017), as well as challenges (Kamal, 2015; van Royen, Moons, Geersing, & van Smeden, 2022), and successful implementations identified in domains beyond those of psychological treatments (e.g., An & Vodovotz, 2021; Hamilton, Rath, Plangger, & Hochmair, 2019; Williams, Feero, Leonard, & Coleman, 2017).

Notwithstanding the limitations mentioned earlier, there has been a recent emergence of evidence supporting the effectiveness of treatment grounded in precision methods (Cohen et al., 2021). A current systematic review and meta-analysis by Nye et al. (2023) identified 17 prospective randomized clinical trials of data-informed/data-driven treatments and found superior outcomes for treatments personalized through precision methods relative to control conditions (6 studies, N = 426) and to standardized treatment (9 studies, N = 5134). In one study (N = 951), algorithm-based treatment recommendations of low-versus high-intensity psychological treatments led to a higher rate of reliable and clinically significant improvement (52.3% vs 45.1%) compared to allocation-as-usual (Delgadillo et al., 2022). Another study found evidence that patients treated with their optimal treatment strategy (problem-solving, motivation-oriented, or mix of both strategies) experienced better outcomes than those treated with a non-optimal strategy (Lutz, Deisenhofer, et al., 2022). In another prospective trial (not included in Nye and colleagues' review), Constantino et al. (2021) found that matching patients to therapists based on therapists' empirically derived strengths in treating specific patient problems led to improved outcomes, relative to a control condition (case assignment as usual). These example studies reveal the growing potential of precision treatment strategies for mental health. The successful implementation of precision methods is vital for translating this potential into practical applications.

The present article aims to [1] summarize evidence for and challenges with implementing existing precision methods; [2] provide a roadmap to help researchers advance the science of personalization; and [3] support clinicians to integrate this growing field of research into psychological practice.

## 1. The implementing precision methods framework

In order to convey the various components involved in the implementation process, we propose an organizing and conceptual framework – the Implementing Precision Methods (IPM) Framework – with a focus on those components that are especially relevant to psychological treatments.[1] The model is based on translational frameworks from implementation science, as well as challenges and successful implementations identified in domains beyond psychological treatments (e.g., An & Vodovotz, 2021; Chambers et al., 2016; Chanfreau-Coffinier et al., 2019; Ward & Ginsburg, 2017). The framework's components are organized into four domains, which we outline below (see Fig. 1), to guide the multifaceted endeavor of implementation and the structure of the article.

Adopting Fig. 1 as a roadmap and guide for the current article, we start at the left focusing on the core of the implementation process, namely with the *clinical and practical applications* (the first domain), which require the consent and cooperation of clients and clinicians. In precision mental health care, these practicalities can quickly interface with *technical aspects* (the second domain) such as the software and hardware used and how it needs to be selected with user-friendliness in mind. Moving beyond the specific technical components, the third domain that plays a role in the implementation process comprises the *statistics* and analytical procedures used in precision mental health care. The robustness of algorithms plays a central role, as does the question of how algorithms can be repeatedly updated in everyday clinical practice in this crucial area. The outer-most domain includes *contextual factors* (the fourth domain) that must be considered, such as legal barriers, funding, societal pressures, clinical policies/guidelines, and the setting in which implementation will occur. In the following sections, the obstacles and opportunities related to each of these domains are examined in greater detail.

## 2. Domain: clinical and practical applications

Clinicians and clients, the two main stakeholders, are central to the implementation of data-driven treatment. Any form of implementation is likely to fail if these two groups of people are not willing participants. For this reason, the two groups of people will be examined separately, and possible difficulties discussed.

### 2.1. Clients' needs and burdens

A barrier at the center of implementation of precision mental health care relates to clients' concerns, which may prevent the concrete application of tools that could improve their outcomes (Greenhalgh et al., 2017). Approaches that are considered more personalized to the individual needs of patients and clients tend to have greater value ratings (Patel et al., 2020). As such, it would be a mistake not to include clients' needs in the development process. For this reason, in this section we focus on the client, with their needs and possible concerns, and address the aspects that require consideration during implementation.

A key consideration for the successful adoption of data-informed treatment recommendations is whether a client is willing and able to provide necessary data to use it in clinical practice (i.e., provide the input needed to generate a prediction for decision support tools). This

[1] We note that although not explicitly included, research plays an important role across all components specified in the conceptual framework.

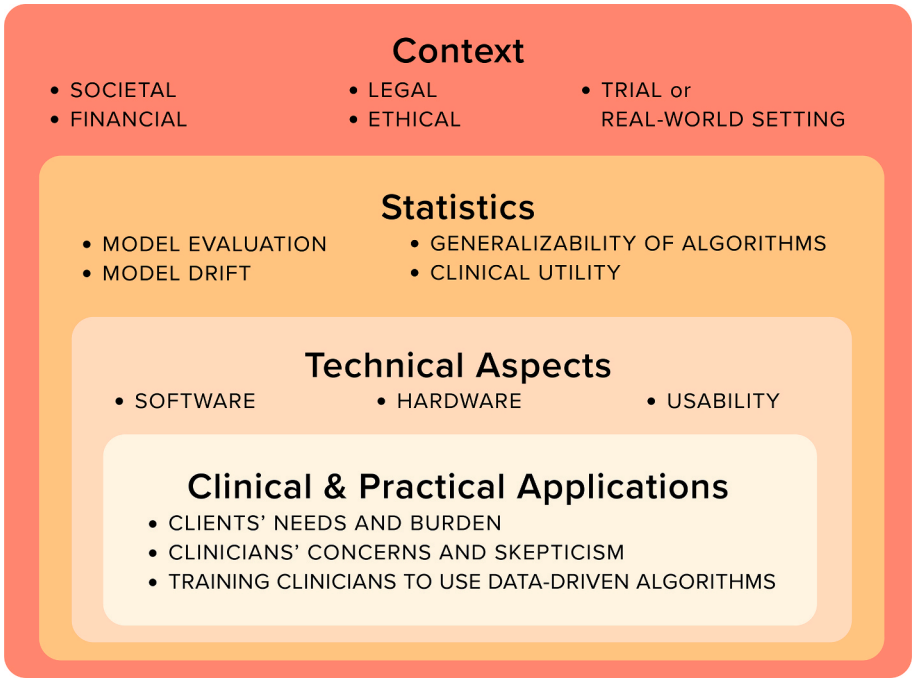


**Fig. 1.** Implementing Precision Methods (IPM) framework
*Note.* This figure illustrates the IPM framework, detailing the implementation components of precision mental healthcare within clinical practice.

relates to the topic of measurement burden that refers to the perceived or actual effort, time, and resources required from individuals to collect, report, or participate in data collection processes. To date, client data primarily comprises information obtained from health care encounters. For example, self- and other-report data, obtained via questionnaires or clinical interviews, are key sources of information in routine care that can be highly informative. While for most clients this type of data provision is straightforward and may provide useful insights for practicing clinicians, some clients may not understand certain instructions or questions due to language barriers or cognitive impairments, and others may simply be too ill for lengthy assessments. Moreover, in child mental health services, multi-informant data (e.g., data provided by the child as well as the parents or teachers) are often considered crucial for making treatment decisions, but acquiring this kind of data can be challenging for some children (Achenbach, McConaughy, & Howell, 1987; De Los Reyes et al., 2015).

Measurement burden increases when *repeated* outcome measurements over the course of treatment are needed to give feedback in addition to data required *prior* to the intervention to match a client to the treatment (Lambert, 2011; Lutz, Schwartz, & Delgadillo, 2022). For example, if an algorithm based on data is to indicate the need for treatment augmentation or for an adaptation of the therapeutic approach in the case that early response is insufficient (Lutz et al., 2017), a session-by-session or even more intensive data sampling scheme would likely be required. This would also be necessary to directly recommend a particular treatment technique or module specific to each session. Obviously, when repeated measurement (for example, intensive longitudinal data from ecological momentary assessment) requires active input from clients (or others, like clinicians or clients' families), the obstacles discussed above (language barriers or cognitive impairments and the like) are multiplied (Bos, Snippe, Bruggeman, Wichers, & van der Krieke, 2019; Shiffman & Stone, 2008). At the same time, evidence suggests that, over time, people become accustomed to the procedures involved in measurement-based care and feedback, and the burden becomes routine (De Jong, Delgadillo, & Barkham, 2023).

Novel approaches like computerized adaptive testing (CAT) – in which relatively few individual items are needed to arrive at a reliable summary assessment of the client – can help address the issue of measurement burden (Gershon, Rothrock, Hanrahan, Bass, & Cella, 2010). In addition, recent advances in Large Language Models (LLMs) like ChatGPT (van Dis, Bollen, Zuidema, van Rooij, & Bockting, 2023) could facilitate passive quantification of relevant data like symptoms and therapy processes from session transcripts. However, so far, these models solely offer the most probable answer or even generate content in the absence of data, a fact not immediately evident to the user. The overarching question remains whether comprehensive safety measures can indeed be established for clinical application, ensuring that symptoms and treatment processes are not merely fabricated or generated by chance. Additionally, as digital technologies and data become an increasing part of everyday life, there are ample data (in addition to self-reported data) that could be used to inform treatment decisions while reducing measurement burden. All sorts of passively collected data (e.g., social media posts, physiological measurements, etc.) can nowadays be assessed rather easily, for example, via mobile phones or wearables. However, digital data may be quite sensitive and ethical concerns may arise (see Context Domain for more details). A key question here is: Are clients willing to share this type of data and allow it to be transmitted, processed, and stored for the purpose of improving their treatment?

Although the issue of data privacy in the context of data-driven decision-support tools in psychiatry and psychology has been surprisingly under-researched in the past, fortunately there have been an increasing number of studies addressing this issue in recent years. For example, studies explored clients' preferences regarding their involvement in data collection and investigated their willingness to share data to facilitate the development of such algorithms (Seltzer et al., 2019). Other studies suggest that clients are open to sharing their personal health data, particularly if those data are used to improve care (Grande, Mitra, Shah, Wan, & Asch, 2013) rather than for-profit research (De Freitas, Silva, Leão Teles, Maia, & Amorim, 2020) or commercial organizations (Aggarwal, Farag, Martin, Ashrafian, & Darzi, 2021). These findings clearly indicate that most clients are willing to share some aspects of their data, particularly when they could be used to enhance their care. However, not all digital data appear to be equally likely to be shared, as

some may be perceived to be more sensitive than others.

It should be emphasized that there are many good reasons why clients may not want to provide their data, even if it could enhance care. For instance, there may be strong concerns about privacy of sensitive information, particularly given the number of high-profile data breaches that have occurred with numerous consumer organizations. In line with this perspective, clients and their families value anonymity and may be worried about misuse of their personal data. For instance, previous genetic-based studies have reported the misuse of private data by third parties as a barrier, particularly if the data were available to employers or insurance companies (Baldwin et al., 2022; Wilde, Meiser, Mitchell, & Schofield, 2010). Clients who identify with a minority status may also be concerned that the predictive algorithm may not be suitable for them or may discriminate against them. These concerns are justified as most existing samples are not representative of diverse populations and Artificial Intelligence (AI) applications can perpetuate biases from historical clinical labels. It is important to acknowledge that bias correction methods exist, but it is still unknown how they might improve the acceptability of precision healthcare in diverse populations (for more details, see Statistics Domain: Generalizability).

In sum, the ability and willingness to share data should both be important considerations for the development and implementation of treatment matching algorithms, as algorithms will fail if they rely on data that many clients are unable or unwilling to provide. Besides this, clients' concerns about loss of agency due to precision treatment algorithms should be addressed through shared decision-making or the explicit inclusion of client preferences during the development process. Additionally, if the process is not communicated in a language comprehensible to clients, concerns regarding comprehension could also be significant. Effective schedule management and enough time spent with clients and their families to explain the process and its therapeutic benefits are often needed (Moreno-Peral et al., 2019). As with all clinical methods that are data-driven, providing clients with information regarding the rationale for working in this particular way is vital in order to establish common ground.

### 2.2. Clinicians' concerns and skepticism

The implementation of recommendations generated by precision methods will occur in the context of real-world clinical practice, which is carried out by clinicians. A key barrier to implementing precision mental health relates to both the overt and covert skepticism felt by many clinicians regarding digital systems and the information they yield (Obermeyer & Weinstein, 2019; Soyster, Song, & Fisher, 2022).

At a basic level, any new routine, whether digital or non-digital, needs to demonstrate to clinicians, who are hard-pressed for time, that the benefits outweigh the investment required for adoption (Baldwin et al., 2022). Despite the large body of research that has arisen in the decades since Meehl's (1954) seminal work reporting actuarial methods to be more reliable than clinician judgements, many clinicians are still hesitant or reluctant to adopt *data-informed* clinical practice. For example, despite measurement-based care being one of the few interventions that has been shown to improve clinical outcomes for psychotherapists (De Jong et al., 2021), a recent study found that only 14% of clinicians used standardized symptom progress measures at least monthly (Jensen-Doss et al., 2018). This may be due, in part, to the problem that data are frequently perceived as either irrelevant or, at worst, as a means of criticizing their clinical work and judgment. A combination of policy driven interventions (Zhu et al., 2021) and other clinician behavior change strategies (Perkins et al., 2007) will likely be needed (see Context Domain: Societal, Financial and Contextual factors). One way to engage clinicians is to evidence the gains for their practice with those clients who are either most challenging or, in their view, most likely to fail to make the gains in therapy that they might be expected to make. Additionally, any such system needs to be perceived by clinicians as augmenting their clinical judgment. Thus, the main challenge is to find ways to create recommendations that enhance the quality of data-informed clinical decisions without undermining a clinicians' expertise or agency. One way of doing this could be to present information from precision treatment rules through a probability and uncertainty framing (e.g., statistical confidence intervals). Furthermore, the information generated by precision algorithms must also be actionable for the provider delivering care; for example, a recommendation that requires a drastic change in a clinician's practice (e.g., learning a new clinical approach) is unlikely to be adopted.

For all the foregoing reasons, it would seem essential to co-design precision method tools with the involvement of clinicians from the beginning of any project in order to satisfy the key objective of ensuring that any yield will be more readily valued by clinicians as having clear gains over existing practices. In short, the rollout of precision methods needs to be aligned with the needs of clinicians and not driven solely by the speed of innovation itself. Involving (scientist-) practitioners as stakeholders in the process of defining, co-designing, and testing prototypes of a developed tool, followed by refinement and evaluation, will increase the likelihood of successful implementation and benefits for clinicians who are more inclined to adopt such systems in their routine practice (Oliveira, Zancul, & Fleury, 2021).

### 2.3. Training clinicians to use data-driven algorithms

In a systematic review on barriers and facilitators of real-world implementation of precision psychiatry, clinician training emerged as a crucial factor, with poor perceived competence in precision methods among staff and insufficient availability of competence and skills training being identified (Baldwin et al., 2022). Given that the utilization of precision methods in mental health care is still in its infancy, most clinicians lack training in various aspects such as interpreting recommendations provided by algorithms and integrating them into clinical decision-making. In this context, a recent study is causing concern by revealing that healthcare professionals exhibit poor proficiency in statistics, frequently coupled with unwarranted self-assurance, even when their responses are incorrect (Lakhlifi, Lejeune, Rouault, Khamassi, & Rohaut, 2023). Therefore, it is essential for clinical training programs to expand and incorporate training on the application of data-informed therapies (Lutz, Deisenhofer, et al., 2022).

It should be noted that different levels of training are required depending on what type of recommendation is made by a system. For example, clear-cut recommendations on what to do with a particular client can be distinguished from those that provide less direct information (e.g., client-specific treatment response trajectories). The more room for interpretation of the data, the more training is necessary to be able to make use of the provided information. Therefore, from a straightforward and time efficient perspective, clear and unambiguous recommendations designed to improve clinical judgment make the most sense (e.g., if the client reports a drop in the alliance greater than *X* points on measure *Y*, implement rupture repair procedures *ABC*). However, research on data-informed treatments is currently far from being able to provide specific recommendations with certainty; providing clinicians with more comprehensive statistical training in concepts relevant to statistical prediction and psychometric feedback (e.g., on measurement error and confidence intervals) will help them to better integrate the interpretation of data into their own clinical knowledge and experience.

Besides increasing clinicians' skills in using data and interpreting algorithm-based feedback, a sound methodological training could also help change therapists' attitudes towards data-informed psychological therapy and their self-efficacy in using such methods. A recent randomized controlled trial on feedback in outpatient psychological therapy identified therapists' attitudes towards feedback and their confidence in using it as relevant predictors of treatment outcome, and the therapist-rated usefulness of the feedback as a moderator of the feedback effect (Lutz, Deisenhofer, et al., 2022). Feedback yielded

improved outcomes compared to the control condition without feedback, when therapists rated it as useful. Therefore, training could increase therapists' willingness and readiness to include data into their decision-making by enabling them to be more confident in using such methods.

To develop the required specialist therapeutic skills, it is helpful to provide hands-on training (and re-training) with practical examples within the specific system used. In addition, to increase clinicians' receptivity to this type of information, it is helpful to highlight the biases introduced by clinical judgment that can be minimized by statistical models (e.g., Aegisdottir et al., 2006; Magnavita & Lilienfeld, 2016; Meehl, 1954; Walfish, McAlister, O'Donnell, & Lambert, 2012). Nevertheless, statistical models can also result in inaccurate predictions due to several factors (e.g., data quality, training sample generalizability) and clinicians need to be aware of the associated implications for the level of certainty with which statistical predictions can be made, especially in regard to minority groups (see domain legal and ethical for more details). In summary, both clinical judgment and statistical predictions may be inaccurate in individual cases. Since it is a process based on probabilities rather than certainties, a reflective professional approach is essential. Furthermore, the rationale of the data-informed approach should be trained comprehensively so that data indicating a lack of treatment progress is not misinterpreted as a criticism of the clinician or their competency, but instead is viewed by clinicians as providing valuable information that can be used to improve outcomes.

Depending on the available financial resources, some kind of research assistant or service evaluator could enhance the training effect by supporting therapists in using data to inform clinical practice. Especially in government-funded services, such an individual could ensure that data-informed methods are fully implemented. Since structural or environmental interventions are often more effective than behavioral interventions, a central monitoring and support entity might be even more effective than just training therapists and appealing to reason. For example, the outpatient clinic at the University of Trier introduced a feedback session to which trainees are invited whenever one of their patients deviates from an expected treatment course. Clinicians are required to participate in these supervision sessions multiple times during their training. During these sessions, the observed patient course, data-based predictions, and recommendations are discussed collaboratively. The focus is on reflecting how to understand the data and how to translate data into practical actions.

Finally, clinical trainees should be taught to adjust and calibrate their clinical judgment, by integrating information from actuarial data and clients' preferences in order to render a coherent and shared treatment decision that is acceptable to the client. This challenge is one of the biggest obstacles for clinical training, since minimal research has been conducted on how to integrate these factors and situations in which these perspectives diverge are of pivotal importance for clinical decision making. Considering the limited research on this topic, we encourage supervisors to make clinicians aware of these gaps and discuss difficult clinical decisions together. More research is needed to enable the design of empirically based training programs in which clinicians can learn the necessary skills to manage these complex tasks.

## 3. Domain: technical aspects

Technical obstacles may impede the implementation of data-informed psychological treatments, such as outdated technological infrastructure or a complex and cumbersome user interface. The following section highlights technical aspects including necessary information on software and hardware as well as considerations regarding the usability of systems.

### 3.1. Software and hardware

A key component within the technical domain comprises both software and hardware used, as well as associated issues of affordability. In the following, we report prerequisites that should be present in a clinic or practice that wants to apply precision mental health care methods. First, a reliable and secure network infrastructure is needed to support data storage and processing for machine learning models (Lutz, 2022). This can include both on-premise servers and cloud-based solutions, depending on the size and scale of the models and the clinic/practice. Second, the necessary hardware (e.g., electronic tablets) and software (e.g., mental health software applications) resources must be available to collect and analyse the data automatically (Gordienko et al., 2021).

In addition to technical equipment, the organizational implementation of continuous high-quality and relevant electronic client data collection is highly relevant (Ohno-Machado, 2018). A distinction can be made between data collected during therapy sessions and that which can be collected outside of sessions. For routine data collection *during treatment sessions,* various hardware devices such as electronic tablets or computers with WIFI connection are available and should be selected depending on the use case and context. For example, tablets are suitable when clients are asked to fill out questionnaires in the waiting room, so that the time in the therapy session can be used efficiently. The relevant data collected may include baseline client characteristics, treatment progress, and other important information. Software applications such as Research Electronic Data Capture (REDCap) can enable mental health data collection and offer automated analysis to facilitate its use in everyday clinical practice (Van Bulck, Wampers, & Moons, 2022). Simple systems can provide total scores and severity categories of depression and anxiety. This information can be used for tracking changes in symptom severity and treatment trajectories over time that provide feedback both to clinicians and clients.

In addition to collecting data at the point of care, some clinics may also use offline portals (electronic health record software e.g., Epic) or online recommender systems (e.g., REDCap) to gather information from clients outside of their treatment sessions. These data may include both active and passive data, collected through a variety of digital means, such as smartphones, tablets, or personal computers. One common method of gathering active *out-of-session data* is using apps or web portals that are connected to a server. These tools can be useful in facilitating longitudinal data collection of ecological momentary assessments from clients including self-report information as symptom severity ratings or other psychometric data (Hamaker & Wichers, 2017), as well as objective cognitive assessments (e.g., TestMyBrain, Germine et al., 2012; Lee et al., 2023). In addition to active data collection, these tools may also be used to gather passive data on granular and objective treatment engagement, such as the tools and pages viewed during a digital therapy session (Enrique, Palacios, Ryan, & Richards, 2019). This may be particularly useful for clinicians to identify therapeutic ingredients of successful treatment for further personalization (Cohen, Barnes-Horowitz, Forbes, & Craske, 2023). Methods for gathering passive data can extend to smartphone and wearable sensors that can continuously and automatically capture clients' behaviors, movements, and physiology in daily life (e.g., social contact via SMS, calls, conversations detected via microphone, social media app use, geolocation; Heinz, Price, Song, Bhattacharya, & Jacobson, 2023; Jacobson & Chung, 2020; Ren et al., 2022).

Lastly, for transferring scientific findings from predictive models into clinical practice, the client data can be integrated into clinical decision support systems (CDSS; Lutz, Schwartz, & Delgadillo, 2022; Schaffrath, Weinmann-Lutz, & Lutz, 2022). These software systems are embedded with clinical data and scientific knowledge that can be used to assist

healthcare providers and individual clinicians in making treatment decisions. Specifically, CDSSs can provide valuable insights and data-informed recommendations that can help clinicians to diagnose and personalize treatment plans for improving client outcomes. Henshall et al. (2017) tested the feasibility of a CDSS on focus groups of clinicians, clients, and caregivers, and found such a tool to be useful for facilitating clinical decision support as well as promoting clinician-client collaboration and client-centered care (Henshall et al., 2017). One example of a comprehensive treatment selection and tracking system, which is also augmented by tools from e-mental health, is the Trier Treatment Navigator (TTN; Lutz, Rubel, Schwartz, Schilling, & Deisenhofer, 2019). This software system includes a data-informed individualized treatment recommendation at the beginning as well as adjustments based on continuous outcome assessments during treatment (Schaffrath et al., 2022). It has been prospectively evaluated and an open-source option for direct use in clinical practice is planned.

By providing information in a timely and easily accessible manner, CDSSs can help to save clinicians' time and improve the precision of clinical judgment. Despite these positive aspects, the cost and affordability of these tools and systems is an important consideration for users. One of the main challenges with implementing machine learning in clinics is the need for specialized hardware and software. These tools can be expensive to purchase and maintain and may require specialized training to use effectively (see clinical training above). Additionally, the cost of collecting and storing the data used by machine learning algorithms can be significant, particularly if these data need to be gathered from many clients. Most clinicians are unlikely to themselves have the combination of resources, capacity, and expertise needed to gather their own data and develop their own precision treatment models. Users may need to carefully evaluate the available possibilities, consider different hardware and software options, and choose the most cost-effective tools and systems. Additionally, it may be necessary to prioritize the data that are collected, focusing on information that is most relevant and valuable for informing treatment decisions (Rost, Binder, & Brückl, 2022). The costs incurred can be seen as a further obstacle, so financing and funding is another important aspect that needs to be considered during implementation (see Domain Societal, Financial, and Contextual Factors). The complexity of these considerations highlights the importance of team-based approaches, wherein different specialists (e.g., clinical-scientists, statisticians, technologists, clinicians) come together to develop and implement precision treatment approaches, thus facilitating the transfer from science to practice.

### 3.2. Usability

Usability considerations can vary based on who the user is (e.g., clinicians or clients). In clinical contexts, primary users interact with these interventions directly, secondary users interact with interventions occasionally or through an intermediary, and tertiary users are indirectly impacted by the intervention (e.g., clients' families) or make choices related to their adoption or deployment (e.g., leadership, administrators).

Although usability is a concept that has traditionally been applied to technologies, recent work has started to apply this concept to psychosocial interventions or innovations (Lyon et al., 2021; Munson et al., 2022). Self-report measures like the System Usability Scale (Bangor, Kortum, & Miller, 2008), which has been widely used for technologies, and the Intervention Usability Scale (Lyon et al., 2021), which is adapted for complex psychosocial interventions, can be used to assess perceived usability (i.e., whether the user thinks the intervention is able or fit to be used). Much can also be learned through direct observation methods like cognitive walkthroughs or "think aloud" methods in which the user interacts with the technology for a specific purpose.

A significant barrier to the broad introduction of support systems into clinical practice, to date, has been low acceptance by clinicians. In a study by Bright et al. (2012), clinicians reported several usability related challenges that had led to low adoption, such as efficiency, helpfulness, content of information displayed, user interface, and operational workflows. Furthermore, one of the first studies to prospectively evaluate a clinical decision support system found that clinicians who were randomized to receive treatment strategy recommendations did not follow the recommendation more often than clinicians of a control group who were not provided with a recommendation (Lutz, de Jong, Rubel, Delgadillo, & Castonguay, 2021). The authors surmised that the figures, effect sizes, and brief descriptions of the recommendations may not have been clear and concrete enough to guide the translation of strategies into therapeutic procedures. These results suggest that there is much room for improvement in facilitating clinicians' implementation of precision therapy recommendations. There is a need to explore how to optimize the presentation of these recommendations so as to increase their usability.

## 4. Domain: statistics

Most precision mental health care approaches support clinicians in their decision-making through recommendations derived from predictive algorithms. An *algorithm* refers to a formal set of steps for converting a set of well-defined inputs into an output. For instance, in precision mental health care, a *treatment matching algorithm* typically refers to specific and explicit steps (usually carried out by a computer) that converts well-specified inputs (e.g., predictive features like patient demographics or clinical factors) into a recommendation that an individual should receive one treatment instead of another (e.g., cognitive therapy vs. antidepressant medication; DeRubeis et al., 2014). Even informal treatment matching can feel algorithmic. For example, when clinicians make intuitive judgments that a client would do better in one type of treatment than another, they translate their client's perceived characteristics into a treatment decision. If you look 'under the hood' of most precision treatment algorithms,[2] you will find recommendations that rely on predictive modeling, such as a set of statistical predictions about how an individual would be expected to respond to two or more treatments. Such models are typically trained on one set of data and then validated on distinct test data. "Training" a model often involves generating "parameter weights" that describe the associations between predictive features and outcomes of interest, thus allowing the model to convert client characteristics into treatment predictions.

When implementing precision treatment approaches, a clinician should have confidence in key features of the algorithms on which they rely, such as the reliability of the predictions and the transferability of the models from the dataset in which they were developed to the context in which they will be used. A thorough and successful evaluation of a prediction algorithm is fundamental before one can begin to encourage its use in clinical practice. In what follows, we will explore the key statistical/methodological questions on the road to real-world clinical impact, including: (1) Is it possible to develop a model that predicts the outcome of interest?; (2) Will updating be necessary to account for model drift (decrease in model performance over time due to changes, e.g., population shift)?; (3) Will a model generalize to different contexts (populations/settings)?; and (4) Will a precision treatment system based on the statistical model demonstrate clinical utility when prospectively compared against the current standard-of-care?

[2] Throughout this paper, we interchangeably use "algorithm" and "model" for simplicity, although they hold nuanced differences. An algorithm presents a defined set of rules to solve a specific problem, while a model represents underlying patterns. In our context, a "predictive algorithm" generates insights based on a model's learned patterns (i.e., outputs), outputs which in isolation may not directly yield treatment recommendations. Such recommendations could require additional steps, like comparing predictions to select the optimal outcome or suggesting heightened care based on preset criteria.

### 4.1. Model evaluation

For any precision treatment algorithm to be considered for use in real-world settings, the model's performance (e.g., accuracy) must be evaluated and determined to be good enough to be helpful (e.g., Efthimiou et al., 2023; Kapelner et al., 2021). For example, if our method is trying to predict depression scores after treatment, the mistakes it makes in these predictions need to be small enough to still be useful for clinical decision-making (Buckmann et al., 2023; Webb et al., 2020). When predicting a binary outcome (e.g., yes/no: such as whether a client's symptoms have improved by the end of therapy), two key metrics are: (1) calibration, which examines how closely the model's predicted likelihood matches the actual outcomes (Buckman et al., 2023; Van Calster, McLernon, van Smeden, Wynants, & Steyerberg, 2019), and (2) discrimination, which assesses the model's ability to differentiate between clients who experienced improvement and those who did not (Steyerberg & Vergouwe, 2014). Unlike simpler models that just predict what will happen to a client after treatment, models that try to predict how effective the treatment will be (like how much a client might benefit from treatment *relative to no treatment*) should be checked for whether they predict these benefits, which is a challenging and active area of research (Kent et al., 2020).

Even though these checks can be used to see how accurate a model is using the data with which it was developed, we need to see how well it generalizes to statistically independent data (i.e., model validation) and there are several ways to achieve this task. First, we have what are called "internal" testing methods such as bootstrapping or k-fold cross-validation (where, for example, the model is developed using 90% of data and tested on the other 10% repeatedly, i.e., a 10-fold cross-validation). These are often used to show that the approach that is being used can predict results based on data from clients the model hasn't seen before (Steyerberg & Harrell, 2016). As long as there is complete independence between the training and test data (and assuming the performance in the test data is not used to adjust the training procedures), the results from these internal testing methods should indicate how well the model would work on new data from the same group of people. However, the results of, for example, 10-fold internal cross-validation come from 10 distinct models, each of which might have picked different factors to consider or given different importance to those chosen factors. Thus, this method does not end up with a single "validated model". Another problem is that researchers often try to optimize their model using these internal tests (for example, by picking the best predicting model from a group of options), which breaks the independence of training and testing data and can result in overfitting. Another option is a train-test split validation, in which, some of the data are kept completely separate while the model is being developed, and the set-aside data are only used to test the model once the model development is completely finished. A drawback of this method is that it is inefficient, as it reduces the available data for model development, as well as for testing, increasing uncertainty about model performance (Steyerberg & Harrell, 2016).

However, using just one set of data to create and test a model can make the model too tailored to that specific data (i.e., overfit) and make it harder for the model to apply to different data sets (Steyerberg & Harrell, 2016). If there are no differences between the group where the predictive model will be used and the group whose data were used to create the model (e.g., if a model was created using data from a large clinic network and then used in the same network), there could be confidence in how well the model is expected to perform. Unfortunately, many real-world implementations will be more complex, involving the use of the model in different situations (e.g., in a different clinic, country, or group of people), or because of situations like changes in the population over time or changes in how care is provided in clinics.

### 4.2. Model drift

Even models that have been adequately evaluated can lose their validity over time (Van Calster, Steyerberg, Wynants, & van Smeden, 2023). This phenomenon is dubbed model drift and refers to the decline of the prediction model's performance because of changes in real-world environments over time, including shifts relating to the impact of, for example, client characteristics, care pathways, and major world events (Davis, Greevy, Walsh, & Matheny, 2020). This can result in predictive models making systematically inaccurate decisions, posing a significant risk for the implementation of these models in routine care. Data in precision mental health are lacking because so few models have been adopted in practice, but evidence from other areas of medicine suggest that drift is a significant threat to model performance and the responsible use of algorithms in healthcare settings (Castro et al., 2022; Davis, Lasko, Chen, Siew, & Matheny, 2017).

While there is, as yet, no unified terminology, the literature often distinguishes between covariate shift, prior probability shift, and concept shift (Moreno-Torres, Reader, Alaiz-Rodíguez, Chawla, & Herrera, 2012). All three refer to changes in the features used to train a model; the first to changes in the distributions of predictors, the second to changes in the distribution of the (categorical) outcome, and the third to changes in the association between predictors and the outcome (e.g., due to a changing context). In the case of an algorithm recommending first-line treatments for depression, for example, a model might include as predictors: time spent at home, the type of treatment received, and age. The distribution of each variable can change systematically for many reasons. There might be rapid changes in the external world (e.g., a global change in time spent at home due to the COVID-19 pandemic) which alter the way therapy is provided. New treatments could become available, modifying which treatments clinicians provide. Alternatively, there could be changes that arise because key characteristics of the sample change over time (e.g., populations using the algorithm might become systemically older).

Another phenomenon is called *label drift* and refers to changes in the distribution of the outcome variable. In the example of precision mental health care for depression, the outcome is primarily the acute phase response or remission. If depression scores rise on average, then the calibration of the model and its evaluation metric (e.g., sensitivity and specificity) will be altered, even though the association between individual predictors and outcome may stay constant. Like all the forms of drift described above, changes in the label can be rapid, slow, or can even occur cyclically – for example, the rise and fall of depressive symptoms with the seasons (Rosenthal et al., 1984).

To keep models accurate, drift needs to be monitored. The population stability index (PSI) is a metric that quantifies how much the distribution of variables has changed between two samples (Yurdakul & Naranjo, 2020). A general convention is that values less than 0.10 mean a "little change", those between 0.10 and 0.25 mean a "moderate change", and those above 0.25 mean a "significant change, action required". PSI has the advantage that it can be used to identify risk of drift using readily available data that could prompt targeted and resource intensive retraining of models in production. Where possible though, continuous, and repeated evaluation of the model performance is the gold standard and may be critical to the continued efficacy of algorithmic decision tools. As new cases are added, meaningful deterioration in the model performance should, ideally, be detected in near real time by monitoring a range of model performance metrics.

Once model drift has been identified, one solution is re-training the model; for example, the re-estimation of some or all regression coefficients and the addition of more predictors. Another option is to recalibrate the model, for example by updating the model intercept. There is, unfortunately, no one-size-fits-all solution to this option as it depends on the type of model, the size, and temporal characteristics of training and test sets. What seems clear is that identifying the source of drift can assist in determining the best course of action and data-driven

tools are emerging to inform model maintenance (Davis et al., 2020). In all cases, real-time monitoring is desirable but practically challenging. As mental health algorithms transition to clinical settings, a fuller integration of research in healthcare systems will be necessary. This includes gathering new data routinely as part of the clinical implementation, and carrying out ongoing research designed to improve and maintain models in tandem.

### 4.3. Generalizability of algorithms

To address another clinically relevant question (e.g., will a given prediction-algorithm work in a different context), external validation, also called a test of generalizability, is needed (Collins, Reitsma, Altman, & Moons, 2015; Ramspeck, Jager, Dekker, Zoccali, & van Diepen, 2021). Tests of generalizability demonstrate that the algorithm is able to provide accurate and valuable predictions in a set of new clients in different services or geographic regions in which the algorithm was not initially developed (Hastie, Tibshirani, & Friedman, 2009).

If the algorithm is not transportable from one sample to another, or one research paradigm to another, then the algorithm's usefulness is limited to the sample and paradigm in which it was generated. This is especially problematic given that the vast majority of research on precision treatment algorithms (and the datasets with which they were developed) has been conducted in wealthy countries, within academic specialist clinics, and in samples that are lacking in diversity. These algorithms are unlikely to generalize to global health contexts (e.g., the global south) and/or to primary care (especially relevant for depression and other common mental health problems). In the meantime, methods for correcting distortions are available that can mitigate but not eliminate these effects (Dang et al., 2022). These methods underscore the importance of fairness analysis in model selection and transparent reporting about the impact of debiasing interventions. The topic of minorities and social inequalities is further discussed in the legal and ethics domain.

Thus, careful consideration is needed when determining the relevant characteristics and identifying the optimal external sample in which to assess generalizability. When choosing the external sample, a balance is needed between dissimilarity (to test for generalizability to new and different data) and similarity of the two datasets (to facilitate generalizability). For example, for algorithms developed using a sample from a specific controlled trial this might be a dataset from another unrelated controlled trial (e.g., similar intervention or client group), or a large naturalistic sample in which the relevant interventions were available (Kessler & Luedtke, 2021). Ideally, the validation sample matches the clinical target population in which the model is intended to be used as closely as possible. Researchers also need to consider the appropriate sample size for validation; methods are available for its determination (Archer et al., 2021; Riley et al., 2021). The use of large e-clinical healthcare records for algorithm development is an area of increasing interest, and provides opportunities to validate algorithms in client subgroups, but also poses issues such as data quality and systematically missing predictors across routine healthcare datasets (Riley et al., 2016).

### 4.4. Clinical utility

Even if all the relevant evaluations of a predictive model's statistical performance determine it to be accurate and robust, this does not guarantee its clinical utility: that is, whether the use of a precision treatment algorithm based on the model will improve clinical practice or client outcomes, relative to a relevant standard of care (e.g., allocation-as-usual, clinical judgment). Evidence of benefit to clinical practice should be evaluated through controlled trials, randomizing clinicians, clients, and services to active use of the prediction model vs. its absence (Delgadillo et al., 2022; Lutz, Deisenhofer, et al., 2022; Weisz et al., 2012).

## 5. Domain: context

The final domain describes the challenges that can arise within a particular setting when the service policies, culture, and available resources are not well-suited to the adoption of data-informed therapies. For example, within a health system, certain service policies, the imposition of a high caseload, or high administrative times that clinicians must meet may hinder the effective use of data-informed treatments. For this reason, it is crucial to consider contextual factors in implementation, as they provide the framework for application. First, social as well as financial barriers will be highlighted followed by an exploration of legal and ethical considerations. Finally, attention will be paid to the context of application, arguing that it makes a difference whether implementation is planned in the context of a trial or in a real-world context.

### 5.1. Societal, financial, and contextual factors

Of all the areas that impact implementation, context, which includes financial and societal aspects, is probably the most difficult to address. This is because it is the furthest removed from the direct influences of a single action by an individual. Thus, societal, financial, and contextual factors must be considered at all stages of designing and implementing data-informed treatments in mental health. This relates to the development of algorithms to further the implementation and usability through technological tools. In addition, successful implementation is complex, as it requires the participation of various groups of professionals and stakeholders: researchers (e.g., quantitative psychologists, statisticians, data analysts), technologists (e.g., software developers, data security experts), policy makers, funders, administrators, clinicians, and patients. This section is intended to provide an overview of the multiple facets within this domain.

Before implementing precision mental health care methods, it is important to address and align around potential conceptual and pragmatic mismatches between different stakeholders. For example, facilities with different clinicians offering different approaches (e.g., cognitive behavioral therapy by clinician *A* and psychodynamic therapy by clinician *B*) can benefit from treatment selection approaches, where the algorithm's recommendation guides the assignment of clients to specific clinicians based on the recommended approach. However, this approach may not be practical for an individual clinician working in a practice setting and trained in one specific approach. This example illustrates that the respective precision methods must be specifically selected according to the clinical context.

One strategy is to use the development phase of the software and algorithms as an opportunity to collect data on key factors to be able to react to them at an early stage. Clinicians and clients, for example, may prioritize ease of use and meaningful outcomes, such as quality of life. Insurance companies may find it more important to focus on the diagnostic status of a client and consider the associated cost of care. Researchers may seek ever-increasing amounts of data and concentrate on researching change mechanisms that other stakeholders may not prioritize. Service providers and administrators may emphasize performance management and want lean systems to deliver the most cost-efficient service to clients, while politicians and lawyers primarily focus on the legal framework, which is significant for the development and implementation of data-informed psychological therapies. Furthermore, if insurance companies or government health systems will not authorize or pay for assessments required for decision-making, the reach of precision methods is likely to be limited. Additionally, policies related to reimbursement for treatments that are not recommended through a precision method need to be clarified to address potential concerns about provider and client autonomy in decision making.

To maximize effectiveness and accuracy of precision mental health care approaches, the clinical use of these systems must also be an integral part of clinicians' competence, training, and practice (see domain

clinical training above). Importantly, treatment selection and adaptation may involve collaboration between multiple health professionals, so coordination and alignment of care is essential, and training and implementation strategies may need to target providers in multiple roles. Most importantly, in addition to the belief that data-driven treatments improve client outcomes, clinicians need time to conduct the necessary data collection and incorporate the results into treatment. For clients too, there is the question of whether the extra effort involved in filling out questionnaires or other tasks is worthwhile. Therefore, it is important that decisions are made at the health policy level that assign important roles to precision mental health care and then promote its implementation through (financial) support and reinforcement.

This leads to the topic of funding, where the question of who will finance the development, testing, and maintenance of such software systems arises. Depending on the country and region, funding opportunities such as grants or competitive funds from the public or private sector may be available. For instance, in the European Union, the programs Horizon Europe and EU4Health 2021–2027 offer specific calls and actions to fund projects in mental health and precision methods, emphasizing technological innovations (European Commission, 2023). In summary, although initial funding approaches for the development of software systems are available in some countries, additional funding will be needed to support the scalable implementation and continued use of these approaches in clinical practice.

### 5.2. Legal and ethical

Implementation of precision mental health care methods in routine clinical practice requires careful consideration of legal and ethical implications and considerations. Federal laws that address clients' rights and protections relevant to data-informed treatment include the Common Rule, HIPAA Privacy and Security Rules, Certificates of Confidentiality, Genetic Information Nondiscrimination Act (GINA), and a complex network of inconsistent and sometimes contradictory laws across differing national states or geographical regions. These laws contain requirements for informed consent, data protection, record creation and retention, and mechanisms for enforcement against data misuse (Hammack, Brelsford, & Beskow, 2019). To further complicate matters, existing frameworks also do not readily map on to precision methods like algorithmic outputs or digital phenotyping (Martinez-Martin, Insel, Dagum, Greely, & Cho, 2018). Regular review of regulation updates that both reflect and impact technological advances will be necessary to maintain a viable degree of synchronicity between the law and clinical practice (Vollmer et al., 2020). In addition, the potential success of precision research and the practice supported by that research depends on public confidence in the research enterprise and resulting clinical practices. As the technological revolution in AI unfolds, the increasing reports of its misuse in social media may affect the standing of precision methods. As no legal, regulatory, technical, or other protective frameworks can ever be entirely sufficient, it is incumbent on researchers, institutional review boards, legislators and policymakers, and other stakeholders to earn and maintain the trust of clients and the public by demonstrably attending to these issues in a transparent manner (Hammack et al., 2019).

The use of clinical prediction models carries additional ethical challenges including issues of privacy, cybersecurity, confidentiality, and device dependability (Fusar-Poli et al., 2022). These issues arise most obviously in the context of partnerships with organizations that exist outside the direct control of healthcare system regulatory frameworks, which may include data and technology vendors, pharmaceutical companies, and medical device manufacturers (Aboujaoude, 2019). Misappropriation or inadequate controls on the release of confidential data or of algorithms using publicly available data, such as social media posts, can have deleterious effects on client lives, including bullying, increased insurance premiums, and adverse action by employers because of discovered medical or psychiatric illness (Chancellor, Birnbaum, Caine, Silenzio, & De Choudhury, 2019; Thapa & Camtepe, 2021). Such risks highlight the importance of both informed consent, and attentiveness to the preservation of privacy, especially in cases where data is collected without consent because such collection is exempt from informed consent requirements or oversight from an institutional review board (e.g., routinely collected individual data; Vollmer et al., 2020).

A broader social justice challenge is the potential for precision approaches to exacerbate social inequities. Precision psychiatry models are only as good as the data on which they are trained, and ever-growing research suggests that training datasets can be unrepresentative and biased (e.g., Gichoya, McCoy, Celi, & Ghassemi, 2021; Obermeyer, Powers, Vogeli, & Mullainathan, 2019). This is particularly problematic considering the technological-halo effect (Fusar-Poli et al., 2022), as there may be a tendency to assume algorithm outputs are not impacted by human biases. Precision models based on prejudicial datasets run the risk of furthering inequities, perhaps through biased allocation of minoritized individuals to treatment modalities without any grounded clinical rationale (e.g., medications vs. psychotherapy) or by furthering precise interventions only for a small subset of individuals. Therefore, there is a tremendous need to increase sampling diversity to help address selection biases, improve the generalizability of precision models, and identify issues with future implementation. This work needs to be guided by a clear ethical framework to identify and resolve concerns at the relevant stages of the design to implementation process, and existing frameworks need to be refined and promoted for this purpose (Fusar-Poli et al., 2022). Precision models developed based on inclusive datasets and used thoughtfully can address inequities (e.g., Pierson, Cutler, Leskovec, Mullainathan, & Obermeyer, 2021).

### 5.3. Trial or real-world setting

Finally, considerations can differ when implementation is planned in the context of a study - as in a randomized-controlled trial with high *internal* validity – versus in a real-world context with high *external* validity. In real world settings where there are multiple modes of treatment available or even clinicians with different trainings (e.g., as in the NHS Talking Therapies program from England; Clark, 2018), the selection of treatment isn't necessarily about the individual clinician but goes well beyond and may involve collaboration between a number of health professionals too (e.g., a GP providing medication and a clinician providing some form of psychological therapy - or, in the case of the English program mentioned above, the nationwide NHS), which is also important for settings where there are few/no specialist/medically trained clinicians. Except for use in a trial, this kind of treatment selection might only be suitable for institutions where practitioners offering different approaches are employed and available (as discussed above). In contrast, real-world treatment strategy/technique selection is more appropriate for everyday use, as clinicians are offered an empirically-based recommendation for a strategy that is part of their treatment package. In addition to the use of the TTN reported earlier, an example is provided by Webb and colleagues (Webb et al., 2022) who showed how precision predictions can be built to decide which skill domain would be most therapeutically beneficial for a given client.

One challenge in implementing precision mental health care is that many clients are treated in small clinical practices where the collection of sufficient data for certain types of precision methods is not possible (e.g., predictive algorithms that involve comparison to a "nearest neighbour"; Lutz et al., 2019). Hence, there is a potential risk that promising precision mental health care approaches could be confined to larger treatment centers and may not be widely disseminated. One potential solution is coordinated outcome data collection, for example as started

**Table 1**
Summary of domain-specific practical suggestions within the Implementing Precision Methods (IPM) framework.

| Domain | Subdomain | Practical Suggestions derived from the manuscript |
|---|---|---|
| Clinical and Practical Applications | Clients' Needs and Burdens | • Providing clients with information regarding the rationale is vital to establish common ground<br>• Balancing the collection of a large amount of data on the one hand with more accurate predictions and patient burden on the other<br>• Taking into account shared decision-making and the deliberate inclusion of client preferences during the process of algorithm development |
| | Clinicians' Concerns and Skepticism | • Precision methods (PMs) need to show their value to time-constrained clinicians<br>• Data-driven recommendations should enhance clinical decisions without undermining clinician expertise<br>• Engaging clinician representatives and scientist-practitioners as key stakeholders in the co-design and development of PMs |
| | Training Clinicians to Use Data-Driven Algorithms | • Expanding clinical training programs to include data-informed therapies<br>• Providing clinicians with more comprehensive statistical training in concepts relevant to statistical prediction and psychometric feedback<br>• Emphasize the importance of a reflective professional approach due to the inherent inaccuracies of both clinical judgment and statistical predictions in individual cases |
| Technical Aspects | Software and Hardware | • Need for specialized hardware and software as well as secure network infrastructure to support data storage<br>• Focusing on information that is most relevant and valuable for informing treatment decisions to reduce data storage costs |
| | Usability | • Recognizing the significance of efficiency, helpfulness, information display, user interface, and recommendation clarity<br>• Inclusion of primary, secondary, and tertiary users as those who influence acceptance and use<br>• Aiming for predictive model errors small enough that the models are still useful for clinical decision making |
| Statistics | Model Evaluation | • Models should be accurate with their training data and generalize to real-world complexity and independent data |
| | Model Drift | • Model performance declines due to evolving real-world conditions, necessitating ongoing monitoring and re-training as required |
| | Generalizability of Algorithms | • Conducting fairness analysis during model selection and transparently report the effects of debiasing interventions<br>• Selecting relevant characteristics carefully and the ideal external sample for assessing generalizability |
| | Clinical Utility | • Statistically sound model evaluations may not ensure clinical usefulness |
| Context | Societal, Financial, and Contextual Factors | • Acknowledge the complexity of implementation, as it requires the participation of various groups of professionals and stakeholders<br>• Health policy decisions should prioritize precision mental health care and facilitate its implementation through financial support and reinforcement |
| | Legal and Ethical | • All parties involved should ensure that informed consent and respect for privacy are guaranteed<br>• Researchers, review boards, legislators, and stakeholders must transparently address issues to earn and maintain public trust<br>• Precision models should be developed based on inclusive data sets and used carefully to eliminate inequities |
| | Trial or Real-World Setting | • Smaller clinics or practices could benefit from initiatives for greater coordinated data collection<br>• Implement adaptive platform trials within routine care to merge implementation with ongoing evaluation and improvement of precision mental health care |

in the German 'KODAP[3]' initiative, a cooperation between university outpatient psychological therapy clinics that includes outcome data harmonization to facilitate coordinated research (Margraf et al., 2021). Via such an initiative, in return for contributing anonymized harmonized treatment data, smaller clinics or practices could benefit from larger data collection. For example, the trajectories of change of their own clients could be compared to others in clinics serving similar demographics, and their clients can also participate in larger trials to gain advantages from the evaluation of novel precision methods.

A possible alternative is to establish a direct connection between research and practice. One effective approach to merging implementation with ongoing evaluation and improvement of precision mental health care techniques would be to utilize adaptive platform trials (Angus et al., 2019) implemented within routine care delivery. For example, the 'leapfrog' trial design has been proposed (Blackwell et al., 2019) and recently demonstrated (Blackwell et al., 2022) as a method to consolidate what would normally be a long-drawn-out treatment development process into a single trial infrastructure and thus facilitate more efficient psychological treatment development. Such a trial requires far fewer participants than standard trial designs due to the use of sequential Bayesian analyses to rapidly identify and discard ineffective treatments. Further, new arms can be added into an ongoing trial, enabling responsivity to new research developments as they emerge from ongoing early-phase work, facilitating a close link between basic science and clinical implementation. Finally, the leapfrog design includes a mechanism for replacing the control condition over time, via another treatment arm hitting a superiority criterion, meaning that once started a leapfrog trial provides a framework for continuous treatment development. Hence, the design provides a method for ongoing testing and implementation of treatment precision methods integrated into routine treatment delivery. We cannot assume that the relative efficacy of any one treatment precision method will generalize across settings or over time, and hence continued evaluation during implementation is important so that decision making is based on 'live' as opposed to increasingly historical evidence (Blackwell & Heidenreich, 2021).

## 6. Discussion

The last decade has seen considerable progress in facilitating our scientific and clinical understanding of precision methods in personalizing psychological therapies, and in developing and rolling out tools for early treatment recommendations and dynamic prediction systems to support ongoing treatments. This paper has presented a comprehensive exploration of implementation within the novel conceptual IPM framework, while Table 1 provides a concise summary of the key domain-specific practical guidelines that also provides an informative overview. If readers are seeking an accessible gateway into the realms of precision methods that they can initiate without engaging in many of the complexities described in this article, then the combination of progress monitoring and session feedback is a credible candidate. These procedures have been documented in academic articles (e.g., see Barkham, De Jong, Delgadillo, & Lutz, 2023), a text specifically written for

[3] Koordination der Datenerhebung und -auswertung an Forschungs-, Lehr- und Ausbildungsambulanzen für Psychotherapie (KODAP).

practitioners (see De Jong et al., 2023), and a series of podcasts.[4] Such procedures can act as a starting point whereby practitioners can utilize data to modify and adjust individual therapy practice and thereby enhance the overall quality of their outcomes.

When considering the field as a whole, one potential barrier we wish to highlight is that precision methods developed in the isolation of "the academic ivory tower" could risk leaving clinicians with the impression that precision methods are overly technical and too far removed from everyday practical work. Thus, researchers need to improve their communication to the public (and not just academic journals) as well as the extent to which they integrate clinicians and other stakeholders throughout the scientific process. User-centered design strategies will be key, such as involving both patients and clinicians in a process of codesign when developing precision tools (Dopp, Parisi, Munson, & Lyon, 2020). The field also needs to prioritize education and, as is usually the case, invest in the training process to capture a new generation of clinicians who understand precision methods as part of their professional concept.

Additionally, the reality is that precision methods may not yield benefits for all clients (Kaiser et al., 2022). Guidance as to when to override such algorithms in practice needs to be considered so as not to give free license to circumnavigate empirically-based decisions. In practical settings, procedures need to be pragmatic and clear. In terms of such guidance, inclusion in clinical guidelines would be a major step forward, but it is probably too early for this to be a realistic option. Ultimately, however, if precision methods are to become mainstream, then new standards for evidence of effectiveness will have to be developed, as has been done for other interventions.

In terms of the practical development of precision treatment systems and the statistical models on which they are based, two possible pathways exist. One option would be to build numerous local, independent precision treatment systems that are specifically tailored to the problems on site. Another possibility would be to develop large independent systems that can be applied everywhere, but which may not cover every need. In terms of broad application, the latter could be an option at the outset. Nevertheless, an alternative perspective is to consider smaller local systems that are integrated within larger ones, as exemplified by the German KODAP initiative mentioned previously. The possibility of direct involvement by local clinicians (e.g., in co-design) that smaller models could allow might increase clinicians' willingness to engage with precision systems. However, this approach poses its own set of significant challenges, particularly with regard to privacy concerns. Ultimately, developing precision treatment systems that can accommodate the diverse needs of stakeholders while maintaining client privacy will be crucial to ensuring successful healthcare interventions.

After the development of precision treatment systems, organizations will need to consider how to manage routine updates and technical upkeep, and identify which individuals or entities will undertake this responsibility. Issues described in the statistics domain (e.g., model drift) suggest an iterative process to keep the models up-to-date and enable them to improve. The consequence is that there will be further costs that will have to be borne and will limit widespread/open access to these precision methods. However, if they are adopted by companies that charge money for their use, precision systems might become impractical for many service providers due to capital costs or a lack of ongoing technical expertise to support such systems. Accordingly, there need to be discussions as to how to ensure that the requirements of such systems do not become a barrier and generate a two-tier system of care provision.

[4] https://www.youtube.com/playlist?list=PLwS02kBxHd2HwzeRkANo5O212U0BM30yG.

## 7. Conclusion

In conclusion, precision mental health methods have the potential to revolutionize the field of clinical psychology by allowing for data-informed and data-driven psychological therapies that have the capability to improve client outcomes. While the research literature on precision methods continues to grow, their application in clinical practice is limited. It is crucial to address the barriers related to dissemination and implementation to ensure that precision methods can benefit clients in real-world clinical settings. This article highlights the importance of considering various factors, including clinical and practical aspects, technical considerations, statistical methods, and contextual frameworks, to facilitate the implementation of precision psychological therapies in mental health contexts. The scope of this article, in combination with the IPM framework with the various domains presented, makes clear how complex the topic of implementation is. Nonetheless, we want to encourage users to push forward with implementation and advocate for pragmatic approaches.

## Funding source declaration

Anne-Katharina Deisenhofer, Julian A. Rubel, Brian Schwartz and Wolfgang Lutz were supported by the German Research Foundation (DFG) under Grant Nr. LU 660/19-1 (504507043) and LU 660/16-1 (493169211).

Christopher Beevers was supported in part by funding from the National Institute of Mental Health (R01MH131750).

Isabel Berwian's work on "Precision psychiatry or treatment selection in depression" is supported by Wellcome Leap as part of the Multi-Channel Psych Program.

Claire Cusack was supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2021320143 (CEC). The content of this manuscript does not necessarily represent the official views of the National Science Foundation.

A grant from the Dutch Research Council (NWO; 016.Veni.195.215 6806) provided financial support for Ellen Driessen's contributions to the preparation of this article.

Toshi A. Furukawa reports personal fees from Boehringer-Ingelheim, DT Axis, Kyoto University Original, Shionogi, SONY and UpToDate, and a grant from Shionogi, outside the submitted work; In addition, TAF has patents 2020–548587 and 2022–082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. And no external funding used for this research.

Nicholas C. Jacobson was partially funded by the National Institute of Mental Health and the National Institute of General Medical Sciences under grant 1 R01 MH123482-01. This work was also supported by an institutional grant from the National Institute on Drug Abuse under grant NIDA-5P30DA02992610.

Lorenzo Lorenzo-Luaces has received funding from the Indiana Clinical and Translational Sciences Institute (CTSI) KL2 Program (Grant: KL2TR002530, PI: B. Tucker Edmonds, PI, institution: NCATS; Grant: UL1TR002529, PIs: S. Moe and S. Wiehe, co-PIs, institution: NACTS).

Vikram Patel was supported by NIMH R01 for a precision medicine study of depression treatment in primary care in India (1R01MH121632-01A1).

Marilyn L. Piccirillo was supported by a grant from National Institutes of Health, AA029459.

Jessica Schleider has received funding for her research from the National Institutes of Health Office of the Director (DP5OD028123), National Institute of Mental Health (R43MH128075), National Science Foundation (2141710), Health Research and Services Association (U3NHP45406-01-00), the Society for Clinical Child and Adolescent Psychology, Hopelab, the Upswing Fund for Adolescent Mental Health, and the Klingenstein Third Generation Foundation. Preparation of this article was supported in part by the Implementation Research Institute (IRI), at the George Warren Brown School of Social Work, Washington

University in St. Louis; through an award from the National Institute of Mental Health (R25MH080916; JLS is an IRI Fellow).

Greg J. Siegle is supported by R01 AT011267, MH074807, MH106591-01, MH096334.

Dr Rudolf Uher is supported by the Canada Research Chairs Program.

Dr. Webb was partially supported by NIMH R01MH116969, NCCIH R01AT011002, the Tommy Fuss Fund and a Young Investigator Grant from the Brain & Behavior Research Foundation.

## CRediT authorship contribution statement

**Anne-Katharina Deisenhofer:** Conceptualization, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Michael Barkham:** Conceptualization, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Esther T. Beierl:** Writing – original draft, Writing – review & editing. **Brian Schwartz:** Conceptualization, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Katie Aafjes-van Doorn:** Writing – original draft, Writing – review & editing. **Christopher G. Beevers:** Writing – original draft, Writing – review & editing. **Isabel M. Berwian:** Writing – original draft, Writing – review & editing. **Simon E. Blackwell:** Writing – original draft, Writing – review & editing. **Claudi L. Bockting:** Writing – original draft, Writing – review & editing. **Eva-Lotta Brakemeier:** Writing – original draft, Writing – review & editing. **Gary Brown:** Writing – original draft, Writing – review & editing. **Joshua E.J. Buckman:** Writing – original draft, Writing – review & editing. **Louis G. Castonguay:** Writing – original draft, Writing – review & editing. **Claire E. Cusack:** Writing – original draft, Writing – review & editing. **Tim Dalgleish:** Writing – original draft, Writing – review & editing. **Kim de Jong:** Writing – original draft, Writing – review & editing. **Jaime Delgadillo:** Writing – original draft, Writing – review & editing. **Robert J. DeRubeis:** Writing – original draft, Writing – review & editing. **Ellen Driessen:** Writing – original draft, Writing – review & editing. **Jill Ehrenreich-May:** Writing – original draft, Writing – review & editing. **Aaron J. Fisher:** Writing – original draft, Writing – review & editing. **Eiko I. Fried:** Writing – original draft, Writing – review & editing. **Jessica Fritz:** Writing – original draft, Writing – review & editing. **Toshi A. Furukawa:** Writing – original draft, Writing – review & editing. **Claire M. Gillan:** Writing – original draft, Writing – review & editing. **J.M. Gómez Penedo:** Writing – original draft, Writing – review & editing. **Peter F. Hitchcock:** Writing – original draft, Writing – review & editing. **Stefan G. Hofmann:** Writing – original draft, Writing – review & editing. **Steven D. Hollon:** Writing – original draft, Writing – review & editing. **Nicholas C. Jacobson:** Writing – original draft, Writing – review & editing. **Daniel R. Karlin:** Writing – original draft, Writing – review & editing. **Chi Tak Lee:** Writing – original draft, Writing – review & editing. **Cheri A. Levinson:** Writing – original draft, Writing – review & editing. **Lorenzo Lorenzo-Luaces:** Writing – original draft, Writing – review & editing. **Riley McDanal:** Writing – original draft, Writing – review & editing. **Danilo Moggia:** Writing – original draft, Writing – review & editing. **Mei Yi Ng:** Writing – original draft, Writing – review & editing. **Lesley A. Norris:** Writing – original draft, Writing – review & editing. **Vikram Patel:** Writing – original draft, Writing – review & editing. **Marilyn L. Piccirillo:** Writing – original draft, Writing – review & editing. **Stephen Pilling:** Writing – original draft, Writing – review & editing. **Julian A. Rubel:** Writing – original draft, Writing – review & editing. **Gonzalo Salazar-de-Pablo:** Writing – original draft, Writing – review & editing. **Jessica L. Schleider:** Writing – original draft, Writing – review & editing. **Paula P. Schnurr:** Writing – original draft, Writing – review & editing. **Stephen M. Schueller:** Writing – original draft, Writing – review & editing. **Greg J. Siegle:** Writing – original draft, Writing – review & editing. **Rudolf Uher:** Writing – original draft, Writing – review & editing. **Ed Watkins:** Writing – original draft, Writing – review & editing. **Christian A. Webb:** Writing – original draft, Writing – review & editing. **Shannon Wiltsey Stirman:** Writing – original draft, Writing – review & editing. **Laure Wynants:** Writing – original draft, Writing – review & editing. **Soo Jeong Youn:** Writing – original draft, Writing – review & editing. **Sigal Zilcha-Mano:** Writing – original draft, Writing – review & editing. **Wolfgang Lutz:** Conceptualization, Project administration, Writing – original draft, Writing – review & editing. **Zachary D. Cohen:** Conceptualization, Project administration, Visualization, Writing – original draft, Writing – review & editing.

## Declaration of competing interest

Michael Barkham has co-produced outcome measures used in routine outcome monitoring and also jointly published a text on routine outcome monitoring but receives no financial payment for these activities.

Christopher G. Beevers has received funding for his research from the National Institutes of Health, Brain and Behavior Foundation, Aiberry Inc., and other private, not-for-profit foundations. He has received income from the Association for Psychological Science for his editorial work and from Orexo, Inc. for serving on a Scientific Advisory Board related to digital therapeutics. Dr. Beevers' financial disclosures have been reviewed and approved by the University of Texas at Austin in accordance with its conflict-of-interest policies.

Tim Dalgleish has received payment for consultancy work with Princeton Biopartners, Boehringer Ingelheim and Felix Inc. He delivers paid workshops on psychological therapies. He receives royalties for books on affect and psychological disorders.

Nicholas C. Jacobson declares the following conflicts of interest. He has edited a book that includes discussion of this topic from which he receives royalties. Additionally, he has accepted paid speaking invitations from various organizations. Dr. Jacobson confirms that these interests do not influence the research, interpretation, or conclusions presented in this manuscript.

Daniel R. Karlin is an employee, officer, and shareholder of MindMed; shareholder and director at Sonara Health; trustee at Trudeau Institute; shareholder and consultant to NightWare; consultant to Otsuka, Tempus, Limitless Ventures, Recovery Delivered, and 4YouandMe.

Cheri A. Levinson has a financial interest in BeWell and Awaken Digital Solutions.

Lorenzo Lorenzo-Luaces has received consulting fees from Syra Health who had no involvement in the current research.

Stephen M. Schueller serves on the Scientific Advisory Board for Headspace for which he receives compensation and has received consulting payments from Boehringer Ingelheim, K Health (Trusst), and Otsuka Pharmaceuticals for unrelated work.

Greg J. Siegle receives royalty payments on a patent regarding a vibroacoustic intervention for psychiatric conditions, licensed to Apollo Neurosciences.

Wolfgang Lutz and Eva-Lotta Brakemeier initiated and chair the Interest Group "Evidence-Based Personalized Psychotherapy Research (ePePsy)" of the German Psychological Association.

## Data availability

No data was used for the research described in the article.

## References

Aboujaoude, E. (2019). Protecting privacy to protect mental health: The new ethical imperative. *Journal of Medical Ethics, 45*(9), 604–607. https://doi.org/10.1136/medethics-2018-105313

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232.

Aegisdottir, S., White, M., Spengler, P., Maugherman, A., Anderson, L., Cook, R., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341–382. https://doi.org/10.1177/0011000005285875

Aggarwal, R., Farag, S., Martin, G., Ashrafian, H., & Darzi, A. (2021). Patient perceptions on data sharing and applying artificial intelligence to health care data: Cross-sectional survey. *Journal of Medical Internet Research, 23*(8), Article e26162. https://doi.org/10.2196/26162

Angus, D. C., Alexander, B. B., Berry, S., Buxton, M., Lewis, R., Paoloni, M., et al. (2019). Adaptive platform trials: Definition, design, conduct and reporting considerations. *Nature Reviews Drug Discovery, 18*(10), 797–807. https://doi.org/10.1038/s41573-019-0034-3

An, G., & Vodovotz, Y. (2021). The rationale and implementation of model-based precision medicine for inflammatory diseases. In Y. Vodovotz, & G. An (Eds.), *Complex systems and computational biology approaches to acute inflammation* (pp. 295–299). Springer. https://doi.org/10.1007/978-3-030-56510-7_16.

Archer, L., Snell, K. I. E., Ensor, J., Hudda, M. T., Collins, G. S., & Riley, R. D. (2021). 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

Baldwin, H., Loebel-Davidsohn, L., Oliver, D., Salazar de Pablo, G., Stahl, D., Riper, H., et al. (2022). Real-world implementation of precision psychiatry: A systematic review of barriers and facilitators. *Brain Sciences, 12*(7), 934. https://doi.org/10.3390/brainsci12070934

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction, 24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Barkham, M. (2023). Smaller effects matter in the psychological therapies: 25 years on from wampold et al. (1997). *Psychotherapy Research, 33*(4), 530–532. https://doi.org/10.1080/10503307.2022.2141589

Barkham, M., De Jong, K., Delgadillo, J., & Lutz, W. (2023). Routine outcome monitoring (ROM) and feedback: Research review and recommendations. *Psychotherapy Research, 33*(7), 841–855. https://doi.org/10.1080/10503307.2023.2181114

Bennett, S. D., & Shafran, R. (2023). Adaptation, personalization and capacity in mental health treatments: A balancing act? *Current Opinion in Psychiatry, 36*(1), 28, 10.1097%2FYCO.0000000000000834.

Blackwell, S. E., & Heidenreich, T. (2021). Cognitive behavior therapy at the crossroads. *International Journal of Cognitive Therapy, 14*(1), 1–22. https://doi.org/10.1007/s41811-021-00104-y

Blackwell, S. E., Schönbrodt, F. D., Woud, M. L., Wannemüller, A., Bektas, B., Rodrigues, M. B., et al. (2022). Demonstration of a 'leapfrog' randomized controlled trial as a method to accelerate the development and optimization of psychological interventions. *Psychological Medicine, 4*, 1–11. https://doi.org/10.1017/S0033291722003294

Blackwell, S. E., Wood, M. L., Margraf, J., & Schönbrodt, F. D. (2019). Introducing the leapfrog design: A simple bayesian adaptive rolling trial design for accelerated treatment development and optimization. *Clinical Psychological Science, 7*(6), 1222–1243. https://doi.org/10.1177/2167702619858071

Bos, F. M., Snippe, E., Bruggeman, R., Wichers, M., & van der Krieke, L. (2019). Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care. *Psychiatric Services, 70*(11), 983–991. https://doi.org/10.1176/appi.ps.201900050

Bright, T. J., Dhurjati, R., Bristow, E., Bastian, L., Coeytaux, R. R., Samsa, G., et al. (2012). Effect of clinical decision-support systems: A systematic review. *Annals of Internal Medicine, 157*(1), 29–43. https://doi.org/10.7326/0003-4819-157-1-201207030-00450

Buckman, J. E. J., Cohen, Z. D., O'Driscoll, C., Fried, E. I., Saunders, R., Ambler, G., et al. (2023). Predicting prognosis for adults with depression using individual symptom data: A comparison of modelling approaches. *Psychological Medicine, 53*(2), 408–418. https://doi.org/10.1017/S0033291721001616

Castro, V. M., Hartl, K. L., Sacks, C. A., Murphy, S. N., Perlis, R. H., & McCoy, T. H., Jr. (2022). Longitudinal validation of an electronic health record delirium prediction model applied at admission in COVID-19 patients. *General Hospital Psychiatry, 74*, 9–17. https://doi.org/10.1016/j.genhosppsych.2021.10.005

Chambers, D. A., Feero, W. G., & Khoury, M. J. (2016). Convergence of implementation science, precision medicine, and the learning health care system. A new model for biomedical research. *JAMA, 315*(18), 1041–1942. https://doi.org/10.1001/jama.2016.3867

Chancellor, S., Birnbaum, M. L., Caine, E. D., Silenzio, V. M., & De Choudhury, M. (2019). A taxonomy of ethical tensions in inferring mental health states from social media. In *In Proceedings of the conference on fairness, accountability, and transparency* (pp. 79–88). https://doi.org/10.1145/3287560.3287587

Chanfreau-Coffinier, C., Peredo, J., Russell, M. M., Yano, E. M., Hamilton, A. B., Lerner, B., et al. (2019). A logic model for precision medicine implementation informed by stakeholder views and implementation science. *Genetics in Medicine, 21*(5), 1139–1154. https://doi.org/10.1038/s41436-018-0315-y

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., et al. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry, 20*(2), 154–170. https://doi.org/10.1002/wps.20882

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

Cohen, Z. D., Barnes-Horowitz, N. M., Forbes, C. N., & Craske, M. G. (2023). Measuring the active elements of cognitive-behavioral therapies. *Behaviour Research and Therapy, 167*. https://doi.org/10.1016/j.brat.2023.104364

Cohen, Z. D., Delgadillo, J., & DeRubeis, R. J. (2021). Personalized treatment approaches. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (7th ed., pp. 673–704). Wiley.

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., DeRubeis, R. J., Hayes, J., Watkins, E., Lewis, G., Byng, R., et al. (2022). The development and internal evaluation of a predictive model to identify for whom mindfulness-based cognitive therapy offers superior relapse prevention for recurrent depression versus maintenance antidepressant medication. *Clinical Psychological Science, 11*(1), 59–76. https://doi.org/10.1177/21677026221076832

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMC Medicine, 13*(1). https://doi.org/10.1186/s12916-014-0241-z

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

Constantino, M. J., Boswell, J. F., Coyne, A. E., Swales, T. P., & Kraus, D. R. (2021). Effect of matching therapists to patients vs assignment as usual on adult psychotherapy outcomes: A randomized clinical trial. *JAMA Psychiatry, 78*(9), 960–969. https://doi.org/10.1001/jamapsychiatry.2021.1221

Dang, V. N., Cascarano, A., Mulder, R. H., Cecil, C., Zuluaga, M. A., Hernández-González, J., et al. (2022). Fairness and bias correction in machine learning for depression prediction: Results from four different study populations. arXiv preprint, arXiv:2211.05321 https://doi.org/10.48550/arXiv.2211.05321.

Davis, S. E., Greevy, R. A., Walsh, C. G., & Matheny, M. E. (2020). Comparison of prediction model performance updating protocols: Using a data-driven testing procedure to guide updating. In *AMIA annual symposium proceedings. AMIA Symposium, 2019* (pp. 1002–1010).

Davis, S. E., Lasko, T. A., Chen, G., Siew, E. D., & Matheny, M. E. (2017). Calibration drift in regression and machine learning models for acute kidney injury. *Journal of the American Medical Informatics Association, 24*(6), 1052–1061. https://doi.org/10.1093/jamia/ocx030

De Freitas, C., Silva, S., Leão Teles, E., Maia, T., & Amorim, M. (2020). Participation in data generation and information sharing: Are patients and carers willing to engage? *The European Journal of Public Health, 30*(Supplement_5). https://doi.org/10.1093/eurpub/ckaa165.370

De Jong, K., Conijn, J. M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review, 85*, Article 102002. https://doi.org/10.1016/j.cpr.2021.102002

De Jong, K., Delgadillo, J., & Barkham, M. (2023). *Routine Outcome Monitoring and feedback in psychological therapies*. Open University Press.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., et al. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*(4), 858–900. https://doi.org/10.1037/a0038498

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., et al. (2022). Stratified care vs stepped care for depression: A cluster randomized clinical trial. *JAMA Psychiatry, 79*(2), 101–108. https://doi.org/10.1001/jamapsychiatry.2021.3539

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., 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), Article e83875. https://doi.org/10.1371/journal.pone.0083875

van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

Dopp, A. R., Parisi, K. E., Munson, S. A., & Lyon, A. R. (2020). Aligning implementation and user-centered design strategies to enhance the impact of health services: Results from a concept mapping study. *Implement Sci Communications, 1*(17). https://doi.org/10.1186/s43058-020-00020-w

Efthimiou, O., Hoogland, J., Debray, T. P. A., Seo, M., Furukawa, T. A., Egger, M., et al. (2023). Measuring the performance of prediction models to personalize treatment choice. *Statistics in Medicine, 42*(8), 118–1206. https://doi.org/10.1002/sim.9665

Enrique, A., Palacios, J. E., Ryan, J. E., & Richards, D. (2019). Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: Secondary analysis of data from a randomized controlled trial. *Journal of Medical Internet Research, 21*(8), Article e12775. https://doi.org/10.2196/12775

European Commission, Directorate-General for Communication. (2023). *The EU in 2022: General report on the activities of the European union*. Publications Office of the European Union. https://data.europa.eu/doi/10.2775/0687.

Fisher, A. J., Bosley, H. G., Fernandez, K. C., Reeves, J. W., Soyster, P. D., Diamond, A. E., et al. (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

Fusar-Poli, P., Koutsouleris, N., Leslie, D., Woopen, C., Calkins, M. E., Dunn, M., et al. (2022). Ethical considerations for precision psychiatry: A roadmap for research and clinical practice. *European Neuropsychopharmacology, 63*(7), 17–34. https://doi.org/10.1016/j.euroneuro.2022.08.001

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

Gershon, R. C., Rothrock, N., Hanrahan, R., Bass, M., & Cella, D. (2010). The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. *Journal of Applied Measurement, 11*(3), 304–314.

Gichoya, J. W., McCoy, L. G., Celi, L. A., & Ghassemi, M. (2021). Equity in essence: A call for operationalising fairness in machine learning for healthcare. *BMJ Health & Care Informatics, 28*(1), Article e100289. https://doi.org/10.1136/bmjhci-2020-100289

Gordienko, Y., Kochura, Y., Taran, V., Gordienko, N., Rokovyi, O., Alienin, O., et al. (2021). "Last mile" optimization of edge computing ecosystem with deep learning models and specialized tensor processing architectures. *Advances in Computers, 122*, 303–341. https://doi.org/10.1016/bs.adcom.2020.10.003. Elsevier.

Grande, D., Mitra, N., Shah, A., Wan, F., & Asch, D. A. (2013). Public preferences about secondary uses of electronic health information. *JAMA Internal Medicine, 173*(19), 1798–1806. https://doi.org/10.1001/jamainternmed.2013.9166

Greenhalgh, T. (2008). Role of routines in collaborative work in healthcare organisations. *BMJ, 337*, a2448. https://doi.org/10.1136/bmj.a2448

Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., A'Court, C., et al. (2017). Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *Journal of Medical Internet Research, 19*(11), e367. https://doi.org/10.2196/jmir.8775

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science, 26*(1), 10–15. https://doi.org/10.1177/0963721416666 65

Hamilton, G., Rath, B., Plangger, A., & Hochmair, M. (2019). Implementation of functional precision medicine for anaplastic lymphoma kinase-rearranged non-small lung cancer. *Precision Cancer Medicine, 2*. https://doi.org/10.21037/pcm.2019.05.03

Hammack, C. M., Brelsford, K. M., & Beskow, L. M. (2019). Thought leader perspectives on participant protections in precision medicine research. *Journal of Law Medicine & Ethics, 47*(1), 134–148. https://doi.org/10.1177/1073110519840493

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer https://hastie.su.domains/Papers/ESLII.pdf.

Hayes, S. C., Hofmann, S. G., Stanton, C. E., Carpenter, J. K., Sanford, B. T., Curtiss, J. E., et al. (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy, 117*, 40–53. https://doi.org/10.1016/j.brat.2018.10.005

Heinz, M. V., Price, G. D., Song, S. H., Bhattacharya, S., & Jacobson, N. C. (2023). Digital biomarkers and passive digital indicators of generalized anxiety disorder. In *Digital mental health: A practitioner's guide* (pp. 13–34). Springer International. https://doi.org/10.1007/978-3-031-10698-9_2.

Henshall, C., Marzano, L., Smith, K., Attenburrow, M. J., Puntis, S., Zlodre, J., et al. (2017). A web-based clinical decision tool to support treatment decision-making in psychiatry: A pilot focus group study with clinicians, patients and carers. *BMC Psychiatry, 17*, 1–10. https://doi.org/10.1186/s12888-017-1406-z

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

Jacobson, N. C., & Chung, Y. J. (2020). Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors, 20*(12), 3572. https://doi.org/10.3390/s20123572

Jensen-Doss, A., Haimes, E. M. B., Smith, A. M., Lyon, A. R., Lewis, C. C., Stanick, C. F., et al. (2018). Monitoring treatment progress and providing feedback is viewed favorably but rarely used in practice. *Administration and Policy in Mental Health and Mental Health Services Research, 45*, 48–61. https://doi.org/10.1007/s10488-016-0763-0

Kaiser, T., Volkmann, C., Volkmann, A., Karyotaki, E., Cuijpers, P., & Brakemeier, E.-L. (2022). Heterogeneity of treatment effects in trials on psychotherapy of depression. *Clinical Psychology: Science and Practice, 29*(3), 294–303. https://doi.org/10.1037/cps0000079

Kamal, M. (2015). Challenges for the clinical implementation of precision medicine trials. In C. Le Turneau, & M. Kamal (Eds.), *Pan-cancer integrative molecular portrait towards a new paradigm in precision medicine*. Springer. https://doi.org/10.1007/978-3-319-22189-2_9.

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*, Article 572532. https://doi.org/10.3389/fdata.2021.572532

Kent, D. M., Van Klaveren, D., Paulus, J. K., D'Agostino, R., Goodman, S., Hayward, R., et al. (2020). The predictive approaches to treatment effect heterogeneity (PATH) statement: Explanation and elaboration. *Annals of Internal Medicine, 172*(1), 35–45. https://doi.org/10.7326/M18-3667

Kessler, R. C., & Luedtke, A. (2021). Pragmatic precision psychiatry - a new direction for optimizing treatment selection. *JAMA Psychiatry, 78*(12), 1384–1390. https://doi.org/10.1001/jamapsychiatry.2021.2500

Lakhlifi, C., Lejeune, F. X., Rouault, M., Khamassi, M., & Rohaut, B. (2023). Illusion of knowledge in statistics among clinicians: Evaluating the alignment between objective accuracy and subjective confidence, an online survey. *Cognitive Research: Principles and Implications, 8*(1), 1–21. https://doi.org/10.1186/s41235-023-00474-1

Lambert, M. J. (2011). What have we learned about treatment failure in empirically supported treatments? Some suggestions for practice. *Cognitive and Behavioral Practice, 18*(3), 413–420. https://doi.org/10.1016/j.cbpra.2011.02.002

Lee, C., Palacios, J., Richards, D., Hanlon, A., Lynch, K., Harty, S., et al. (2023). The Precision in Psychiatry (PIP) study: Testing an internet-based methodology for accelerating research in treatment prediction and personalisation. *BMC Psychiatry, 23*(1), 25. https://doi.org/10.1186/s12888-022-04462-5

Lorenzo-Luaces, L., Peipert, A., De Jesús Romero, R., Rutter, L. A., & Rodriguez-Quintana, N. (2021). Personalized medicine and cognitive behavioral therapies for depression: Small effects, big problems, and bigger data. *International Journal of Cognitive Therapy, 14*(1), 59–85. https://doi.org/10.1007/s41811-020-00094-3

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science, 7*(3), 445–461. https://doi.org/10.1177/21677026188154

Lutz, W. (2022). Data-informed advances and technology augmentation. *Cognitive and Behavioral Practice, 29*(3). https://doi.org/10.1016/j.cbpra.2022.02.008, 591-591.

Lutz, W., Arndt, A., Rubel, J., Berger, T., Schroder, J., Spath, C., et al. (2017). Defining and predicting patterns of early response in a web-based intervention for depression. *Journal of Medical Internet Research, 19*(6), Article e206. https://doi.org/10.2196/jmir.7367

Lutz, W., de Jong, K., Rubel, J., Delgadillo, J., & Castonguay, L. (2021). Measuring, predicting, and tracking change in psychotherapy. In M. Barkham (Ed.), *Bergin and Garfields's handbook of psychotherapy and behavior change* (50th ed., pp. 89–113). Wiley.

Lutz, W., Deisenhofer, A. K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., et al. (2022). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology, 90*(1), 90–106. https://doi.org/10.1037/ccp0000642

Lutz, W., Rubel, J., 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*, Article 103438. https://doi.org/10.1016/j.brat.2019.103438

Lutz, W., Schwartz, B., & Delgadillo, J. (2022). Measurement-based and data-informed psychological therapy. *Annual Review of Clinical Psychology, 18*, 71–98. https://doi.org/10.1146/annurev-clinpsy-071720-014821

Lyon, A. R., Pullmann, M. D., Jacobson, J., Osterhage, K., Al Achkar, M., Renn, B. N., et al. (2021). Assessing the usability of complex psychosocial interventions: The Intervention Usability Scale. *Implementation Research and Practice, 2*. https://doi.org/10.1177/2633489520987828

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.

Margraf, J., Hoyer, J., Fydrich, T., In-Albon, T., Lincoln, T., Lutz, W., et al. (2021). The cooperative revolution reaches clinical psychology and psychotherapy: An example from Germany. *Clinical Psychology in Europe, 3*(1), e4459. https://doi.org/10.32872/cpe.4459

Martinez-Martin, N., Insel, R. R., Dagum, P., Greely, H. T., & Cho, M. K. (2018). Data mining for health: Staking out the ethical territory of digital phenotyping. *Npj Digital Medicine, 1*(68). https://doi.org/10.1038/s41746-018-0075-8

Meehan, A. J., Lewis, S. J., Fazel, S., Fusar-Poli, P., Steyerberg, E. W., Stahl, D., et al. (2022). Clinical prediction models in psychiatry: A systematic review of two decades of progress and challenges. *Molecular Psychiatry, 27*(6), 2700–2708. https://doi.org/10.1038/s41380-022-01528-4

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

Moggia, D., Saxon, D., Lutz, W., Hardy, G. E., & Barkham, M. (2023). Applying precision methods to treatment selection for moderate/severe depression in person-centered experiential therapy or cognitive behavioral therapy. *Psychotherapy Research, 1–16*. https://doi.org/10.1080/10503307.2023.2269297. Advance online publication.

Moreno-Peral, P., Conejo-Cerón, S., Fernández, A., Martín-Peréz, C., Fernández-Alonso, C., Rodríguez-Bayón, A., et al. (2019). Family physicians' views on participating in prevention of major depression. The predictD-EVAL qualitative study. *PLoS One, 14*(5), Article e0217621. https://doi.org/10.1371/journal.pone.0217621

Moreno-Torres, J. G., Reader, T., Alaiz-Rodíguez, R., Chawla, N. V., & Herrera, F. (2012). A unifying view on dataset shift in classification. *Pattern Recognition, 45*(1), 521–530. https://doi.org/10.1016/j.patcog.2011.06.019

Munson, S. A., Friedman, E. C., Osterhage, K., Allred, R., Pullmann, M. D., Areán, P. A., et al., UW ALACRITY Center Researchers. (2022). Usability issues in evidence-based psychosocial interventions and implementation strategies: Cross-project analysis. *Journal of Medical Internet Research, 24*(6), Article e37585. https://doi.org/10.2196/37585

Norcross, J. C., & Cooper, M. (2021). *Personalizing psychotherapy: Assessing and accommodating patient preferences*. American Psychological Association. https://doi.org/10.1037/0000221-000

Nye, A., Delgadillo, J., & Barkham, M. (2023). Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*. https://doi.org/10.1037/ccp0000820. Advance Online Publication.

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), 447–453. https://doi.org/10.1126/science.aax234

Obermeyer, Z., & Weinstein, J. N. (2019). Adoption of artificial intelligence and machine learning is increasing, but irrational exuberance remains. *NEJM Catalyst Innovations in Care Delivery, 1*(1). https://doi.org/10.1056/CAT.19.1090

Ohno-Machado, L. (2018). Data science and artificial intelligence to improve clinical practice and research. *Journal of the American Medical Informatics Association, 25*(10). https://doi.org/10.1093/jamia/ocy136, 1273-1273.

Oliveira, M., Zancul, E., & Fleury, A. L. (2021). Design thinking as an approach for innovation in healthcare: Systematic review and research avenues. *BMJ Innovations, 7*, 491–498. https://doi.org/10.1136/bmjinnov-2020-000428

Patel, S., Akhtar, A., Malins, S., Wright, N., Rowley, E., Young, E., et al. (2020). The acceptability and usability of digital health interventions for adults with depression,

anxiety, and somatoform disorders: Qualitative systematic review and meta-synthesis. *Journal of Medical Internet Research, 22*(7), Article e16228. https://doi.org/10.2196/16228

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

Perkins, M. B., Jensen, P. S., Jaccard, J., Gollwitzer, P., Oettingen, G., Pappadopulos, E., et al. (2007). Applying theory-driven approaches to understanding and modifying clinicians' behavior: What do we know? *Psychiatric Services, 58*(3), 342–348.

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. (2008). *The case formulation approach to cognitive-behavior therapy*. Guilford Press.

Persons, J. B., Koerner, K., Eidelman, P., Thomas, C., & Liu, H. (2016). Increasing psychotherapists' adoption and implementation of the evidence-based practice of progress monitoring. *Behaviour Research and Therapy, 76*, 24–31. https://doi.org/10.1016/j.brat.2015.11.004

Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S., & Obermeyer, Z. (2021). An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nature Medicine, 27*(1), 136–140. https://doi.org/10.1038/s41591-020-01192-7

Ramspeck, C. L., Jager, K. J., Dekker, F. W., Zoccali, C., & van Diepen, M. (2021). External validation of prognostic models: What, why, how, when and where? *Clinical Kidney Journal, 14*(1), 49–58. https://doi.org/10.1093/ckj/sfaa188

Ren, B., Balkind, E. G., Pastro, B., Israel, E. S., Pizzagalli, D. A., Rahimi-Eichi, H., et al. (2022). Predicting states of elevated negative affect in adolescents from smartphone sensors: A novel personalized machine learning approach. *Psychological Medicine*, 1–9. https://doi.org/10.1017/S0033291722002161

Riley, R. D., Debray, T. P., Collins, G. S., Archer, L., Ensor, J., van Smeden, M., et al. (2021). Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Statistics in Medicine, 40*(19), 4230. https://doi.org/10.1002/sim.9025

Riley, R. D., Ensor, J., Snell, K. I. E., Debray, T. P., Altman, D. G., Moons, K. G., et al. (2016). External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges. *BMJ, 353*, i3140. https://doi.org/10.1136/bmj.i3140

Robert, G., Greenhalgh, T., MacFarlane, F., & Peacock, R. (2010). Adopting and assimilating new non-pharmaceutical technologies into health care: A systematic review. *Journal of Health Services Research & Policy, 15*(4), 243–250. https://doi.org/10.1258/jhsrp.2010.009137

Rosenthal, N. E., Sack, D. A., Gillin, J. C., Lewy, A. J., Davenport, Y., Mueller, P., et al. (1984). Seasonal affective disorder: A description of the syndrome and preliminary findings with light therapy. *Archives of General Psychiatry, 41*(1), 72–80. https://doi.org/10.1001/archpsyc.1984.01790120076010

Rost, N., Binder, E. B., & Brückl, T. M. (2022). Predicting treatment outcome in depression: An introduction into current concepts and challenges. *European Archives of Psychiatry and Clinical Neuroscience, 273*(1), 113–127. https://doi.org/10.1007/s00406-022-01418-4

van Royen, F. S., Moons, K. G. M., Geersing, G. J., & van Smeden, M. (2022). Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. *European Respiratory Journal, 60*(3), Article 2200250. https://doi.org/10.1183/13993003.00250-2022

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

Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., et al. (2020). Implementing precision psychiatry: A systematic review of individualized prediction models for clinical practice. *Schizophrenia Bulletin, 47*(2), 284–297. https://doi.org/10.1093/schbul/sbaa120

Schaffrath, J., Weinmann-Lutz, B., & Lutz, W. (2022). The Trier Treatment Navigator (TTN) in action: Clinical case study on data-informed psychological therapy. *Journal of Clinical Psychology, 78*, 2016–2028. https://doi.org/10.1002/jclp.23362

Schwartz, B., Cohen, Z. D., Rubel, J. A., Zimmermann, D., Wittmann, W. W., & Lutz, W. (2021). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research, 31*(1), 33–51. https://doi.org/10.1080/10503307.2020.1769219

Seltzer, E., Goldshear, J., Guntuku, S. C., Grande, D., Asch, D. A., Klinger, E. V., et al. (2019). Patients' willingness to share digital health and non-health data for research: A cross-sectional study. *BMC Medical Informatics and Decision Making, 19*(1), 1–8. https://doi.org/10.1186/s12911-019-0886-9

Shiffman, S., & Stone, A. A. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Soyster, P. D., Song, F., & Fisher, A. J. (2022). Clinician opinions: Perceived utility of and barriers to incorporating pretreatment ecological momentary assessment into clinical care. *Behavior Therapy, 54*(2), 200–213. https://doi.org/10.1016/j.beth.2022.08.005

Steyerberg, E. W., & Harrell, F. E., Jr. (2016). Prediction models need appropriate internal, internal-external, and external validation. *Journal of Clinical Epidemiology, 69*, 247–249. https://doi.org/10.1016/j.jclinepi.2015.04.005

Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal, 35*(29), 1925–1931. https://doi.org/10.1093/eurheartj/ehu207

Szatmari, P., & Susser, E. (2022). Being precise about precision mental health. *JAMA Psychiatry, 79*(12), 1149–1150. https://doi.org/10.1001/jamapsychiatry.2022.3391

Thapa, C., & Camtepe, S. (2021). Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Computers in Biology and Medicine, 129*, Article 104130. https://doi.org/10.1016/j.compbiomed.2020.104130

Tran, A., Klossner, Q., Crain, T., & Prasad, V. (2020). Shifting, overlapping and expanding use of "precision oncology" terminology: A retrospective literature analysis. *BMJ Open, 10*(6), Article e036357. https://doi.org/10.1136/bmjopen-2019-036357

Van Bulck, L., Wampers, M., & Moons, P. (2022). Research electronic data capture (REDCap): Tackling data collection, management, storage, and privacy challenges. *European Journal of Cardiovascular Nursing, 21*(1), 85–91. https://doi.org/10.1093/eurjcn/zvab104

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The achilles heel of predictive analytics. *BMC Medicine, 17*, 230. https://doi.org/10.1186/s12916-019-1466-7

Van Calster, B., Steyerberg, E. W., Wynants, L., & van Smeden, M. (2023). There is no such thing as a validated prediction model. *BMC Medicine, 21*, 70. https://doi.org/10.1186/s12916-023-02779-w

Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., et al. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ, 368*, Article 16927. https://doi.org/10.1136/bmj.l6927

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. Psychological reports. *Psychological Reports, 110*(2), 639–644. https://doi.org/10.2466/02.07.17.PR0.110.2.639-644

Ward, R., & Ginsburg, G. G. (2017). Local and global challenges in the clinical implementation of precision medicine. In G. G. Ginsburg, & H. F. Willard (Eds.), *Genomic and precision medicine* (3rd ed., pp. 105–117). Academic Press. https://doi.org/10.1016/B978-0-12-800681-8.00007-4.

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

Webb, C. A., Forgeard, M., Israel, E. S., Lovell-Smith, N., Beard, C., & Björgvinsson, T. (2022). Personalized prescriptions of therapeutic skills from patient characteristics: An ecological momentary assessment approach. *Journal of Consulting and Clinical Psychology, 90*(1), 51–60. https://doi.org/10.1037/ccp0000555

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., et al., 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

Wilde, A., Meiser, B., Mitchell, P. B., & Schofield, P. R. (2010). Public interest in predictive genetic testing, including direct-to-consumer testing, for susceptibility to major depression: Preliminary findings. *European Journal of Human Genetics, 18*(1), 47–51. https://doi.org/10.1038/ejhg.2009.138

Williams, J. K., Feero, W. G., Leonard, D. G., & Coleman, B. (2017). Implementation science, genomic precision medicine, and improved health: A new path forward? *Nursing Outlook, 65*(1), 36–40. https://doi.org/10.1016/j.outlook.2016.07.014

Yurdakul, B., & Naranjo, J. (2020). Statistical properties of the population stability index. *Journal of Risk Model Validation, 14*(4), 89–100. https://doi.org/10.21314/JRMV.2020.227

Zhu, N. J., Ahmad, R., Holmes, A., Robotham, J. V., Lebcir, R., & Atun, R. (2021). System dynamics modelling to formulate policy interventions to optimise antibiotic prescribing in hospitals. *Journal of the Operational Research Society, 72*(11), 2490–2502. https://doi.org/10.1080/01605682.2020.1796537