



Predicting treatment outcome in depression: an introduction into current concepts and challenges

Nicolas Rost^{1,2} · Elisabeth B. Binder¹ · Tanja M. Brückl¹

Received: 26 November 2021 / Accepted: 11 April 2022 / Published online: 19 May 2022
© The Author(s) 2022

Abstract

Improving response and remission rates in major depressive disorder (MDD) remains an important challenge. Matching patients to the treatment they will most likely respond to should be the ultimate goal. Even though numerous studies have investigated patient-specific indicators of treatment efficacy, no (bio)markers or empirical tests for use in clinical practice have resulted as of now. Therefore, clinical decisions regarding the treatment of MDD still have to be made on the basis of questionnaire- or interview-based assessments and general guidelines without the support of a (laboratory) test. We conducted a narrative review of current approaches to characterize and predict outcome to pharmacological treatments in MDD. We particularly focused on findings from newer computational studies using machine learning and on the resulting implementation into clinical decision support systems. The main issues seem to rest upon the unavailability of robust predictive variables and the lacking application of empirical findings and predictive models in clinical practice. We outline several challenges that need to be tackled on different stages of the translational process, from current concepts and definitions to generalizable prediction models and their successful implementation into digital support systems. By bridging the addressed gaps in translational psychiatric research, advances in data quantity and new technologies may enable the next steps toward precision psychiatry.

Keywords Major depressive disorder · Treatment outcome · Predictive modeling · Clinical decision support system · Precision psychiatry

Introduction

With over 300 million affected people worldwide, depressive disorders have become one of the main causes of disability [1, 2]. Even though there has been an increasing number of studies investigating the optimization of treatment for major depressive disorder (MDD), response rates in patients remain unsatisfactory [3, 4]. In fact, rates have not much improved since the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study reported in 2006 that only 30% of patients reach the clinical goal of remission, i.e., the absence of symptoms, after the first trial of medication

[5]. These numbers need to be taken seriously given the high level of suffering during depressive episodes, the high risk for suicide and comorbidities, and the huge social and economic impact [6, 7]. The question of *what* constitutes the best treatment option for a *specific* patient with a depressive episode under certain individual circumstances is still difficult to answer. Approaches that allow the matching of patients with personalized treatments, often termed ‘precision medicine’, are widely called for in psychiatry [8, 9]. Particularly in early stages of MDD treatment, it is often unclear whether an individual patient will profit most from pharmacotherapy or if other approaches, such as psychotherapy, brain stimulation, or a combination of treatments, might be more beneficial [10]. Models predicting treatment outcome on the basis of individual baseline characteristics can inform the stratification of patients according to their response chances and consequently, the physician’s choice of individualized treatment strategies. In oncology, for example, molecular approaches for tumor characterization have led to the discovery of important subtypes and greatly

✉ Nicolas Rost
nicolas_rost@psych.mpg.de

¹ Department of Translational Research in Psychiatry,
Max Planck Institute of Psychiatry, Kraepelinstraße 2-10,
80804 Munich, Germany

² International Max Planck Research School for Translational
Psychiatry, Munich, Germany

improved individualized treatments [11, 12]. However, in psychiatry, prediction models have not yielded any reliable and valid (bio)markers that are ready for incorporation into clinical tools to support diagnoses or guide treatment decisions (for a review, see [13]). For the treatment of specific psychiatric disorders, such as MDD, mental health professionals can refer to evidence-based, mostly country-specific, guidelines that have been formulated by a committee of experts, such as the American Psychiatric Association [14] or corresponding organization in other countries (e.g., Germany; [15]). These guidelines typically recommend, depending on depression severity, different initial treatment trials as well as a stepwise increase in treatment intensity if initial treatments fail. To some extent, they also take individual patient characteristics into account by adapting treatment recommendations to specific comorbidity or symptom patterns and the patient's prior subjective experience with tolerability and efficacy of certain antidepressants. Standardized approaches in the treatment of MDD, such as guideline- and measurement-based [16] treatments, can help to improve treatment success rates [17]. However, treatment guidelines for MDD are also limited by the non-availability of accurate and validated makers of treatment outcome that are needed for the personalization of treatment. Therefore, treatment administration in MDD is often based on the physician's individual experiences and the patient's personal preferences [18], potentially adding to the low success rates of MDD treatment [19]. With the current lack of personalized treatment, it is more likely that a chosen treatment will be inefficient than efficient for a certain patient [20].

Thus, a better understanding of individual factors contributing to treatment outcome in MDD continues to be a major topic in psychiatry. The present review summarizes definitions of and issues with the current concepts of treatment outcome and provides an introduction into approaches to study and predict antidepressant outcome in MDD. It focuses on clinical implications from these approaches and on implementations into clinical decision support systems.

How is treatment outcome in MDD defined?

In the absence of measurable biological indicators of depression severity, it is important to understand how treatment outcome in MDD is commonly defined and how patients are evaluated based on their rate of recovery.

Changes in symptom severity

In clinical studies, the efficacy of any kind of treatment in MDD as in other psychiatric disorders is routinely assessed with symptom questionnaires, including both clinician-based ratings as well as patient self-ratings. Table 1

summarizes the most typical definitions of treatment outcome based on these ratings. Among the most commonly used scales are the Hamilton Rating Scale for Depression (HDRS; [21]), the Montgomery–Åsberg Depression Rating Scale (MADRS; [22]), the Quick Inventory of Depressive Symptomatology (QIDS; [23]), and the Beck Depression Inventory (BDI; [24]). While the HDRS and MADRS are both clinician-based ratings and require a certain amount of clinical training from the rater [25, 26], the QIDS and BDI are scales based on self-assessments. Even though all these scales were initially created to measure the same construct, i.e., MDD symptom severity, studies have shown that they are not entirely congruent but should rather be used as complementary measures, irrespective of their assessment method [27, 28].

Symptom questionnaires are commonly analyzed by adding up their single items into a sum score. Treatment outcome can then be evaluated by simply interpreting this sum score after a certain length of treatment or by comparing it to a baseline score. However, even though the scales are semiquantitative, binary outcome definitions are widely used, the most common ones being 'response' and 'remission'. Treatment response implies a reduction of symptom severity compared to baseline severity by a certain amount (usually by at least 50%), whereas remission requires symptom scores to drop below a certain threshold (e.g., ≤ 7 on the 17-item HDRS; [29]). Since the concept of response relies on the percentage change in symptom severity, it strongly depends on the baseline score. Remission, on the contrary, does not rely on baseline symptom severity at all. From a clinical perspective, remission is the more desired outcome as remitted patients are generally considered symptom-free and, for the time being, fully recovered. Compared to patients who report residual symptoms after treatment (e.g., response without remission), remitters show a reduced risk of subsequent relapse [30, 31].

If depressive symptoms are continuously measured over time, outcome definitions are not restricted to absolute or relative measures, such as response or remission. Instead, trajectories of symptom development over time can be considered to evaluate treatment success. Many longitudinal studies and clinical trials collect data by applying symptom scales on a weekly basis, which allows outcome definitions built on data from more than one or two timepoints. With this information, more refined interpretations of treatment effects can be made for individual patients. Furthermore, symptom trajectories can be used to identify subgroups of patients with similar outcome patterns but different dynamics in change. With increases in computing power, advances in statistical methods and sufficient sample sizes, such

Table 1 Definitions of treatment outcome in MDD

Concept	Operationalization in studies	Evaluation
<i>Symptom severity</i>	Raw (sum) score derived from a depression severity questionnaire	Continuous measurement Independent from baseline severity As a stand-alone measure (without reference to baseline measure) no clear clinical interpretation
<i>Change in symptom severity</i>	Percentage change or difference of (sum) scores derived from depression severity questionnaire between two time points	Continuous measurement Highly dependent on baseline value
<i>Partial response</i>	Reduction of (sum) score on a depression severity questionnaire usually by 25–49% between treatment start and a specific treatment week	Dichotomous outcome measure (yes/no) Dichotomization by arbitrary threshold Early indicator for stable response later Highly dependent on baseline value
<i>Response</i>	Reduction of (sum) score on a depression severity questionnaire usually by at least 50% between treatment start and a specific treatment week	Dichotomous outcome measure (yes/no) Dichotomization by arbitrary threshold Highly dependent on baseline value Does not necessarily imply that core symptoms have improved
<i>Remission</i>	Depression severity score below a certain threshold. Cut-off values vary by scale: HDRS-17: ≤ 8 / ≤ 7 (e.g., [5, 15]) HDRS-21: ≤ 8 [18] MADRS: ≤ 6 [15] QIDS-SR: ≤ 5 [5] BDI-II: ≤ 9 [5]	Dichotomous outcome measure (yes/no) Dichotomization by arbitrary threshold Reflects the treatment goal Compares an almost symptom-free group to the remainder group
<i>Symptom trajectories</i>	Identification of patient subgroups with distinct patterns of change in symptom severity over duration of treatment; patients with similar pattern are assigned to the same group; methods often based on unsupervised machine learning	Categorical outcome measure Data-driven method of outcome definition No dichotomization or cut-off value needed Dependent on selected variables and scales Heterogeneous methods and algorithms for subgroup identification
<i>Treatment resistance</i>	Often defined as no significant symptom reduction after at least two adequate antidepressant trials coming from different pharmacological classes; staging approaches	Heterogeneous definitions and staging models (categorical vs. dimensional approaches) Potentially stigmatizing terminology
<i>Functional recovery</i>	Changes in quality of life and disability measures	Important additional measures that imply a broader understanding of recovery (beyond symptom reduction) Often not assessed in clinical studies Subscales partly not applicable in inpatient settings
<i>Recovery in daily functioning beyond symptom reduction</i>		

HDRS Hamilton Rating Scale for Depression, *MADRS* Montgomery–Åsberg Depression Rating Scale, *QIDS-SR* Quick Inventory of Depressive Symptomatology (Self-Report), *BDI-II* Beck Depression Inventory-II

approaches are becoming more and more prevalent [32–36].

Treatment resistance

In contrast to response and remission, non-response and non-remission can be precursors of so-called ‘treatment-resistant depression’ (TRD). Definitions of TRD also depend primarily on scores from symptom questionnaires and are mainly focusing on pharmacotherapy. Even though there is no unique definition [37], TRD is most commonly described as a major depressive episode with no response after two or more trials of adequate antidepressant medication coming from different pharmacological classes [38–40]. Still, although this definition seems to be the most prevalent and a useful common ground, many different definitions exist. Some of them vary fundamentally in their criteria, making them difficult to compare [38, 41].

Recovery of cognition and daily functioning

Apart from reduction of symptom severity and failed treatment trials, the desired outcome after a depressive episode also includes other aspects of the patient’s recovery. Ideally, patients return to the same (or even a higher) level of well-being as well as to their way of living from before the disorder, including their daily functioning, i.e., their work, social contacts, and general quality of life [42, 43]. This overarching goal of MDD treatment, helping patients to achieve all aspects of recovery, seems to be a stepwise process. For patients with acute moderate or severe episodes, a reduction of symptoms is naturally the first target. Hence, in clinical studies, especially in inpatient settings, symptom severity is more commonly measured than levels of functioning and positive affect [44], the assessments of which are not necessarily well-suited for routine use [19].

Nevertheless, restoration of daily functioning and positive affect are important factors of a holistic picture of recovery. Any potentially impaired cognitive abilities, such as attention, learning, memory, and executive functions [45], should improve, as should components of positive affect, such as optimism and self-confidence [46]. Whereas cognition is routinely assessed using different neurocognitive tests or batteries [47], functional aspects are less well defined [19]. Still, numerous scales and questionnaires with varying foci exist, including the Global Assessment of Functioning [48], the Quality of Life Enjoyment and Satisfaction Questionnaire [49], and the World Health Organization Disability Assessment Schedule [50].

Prediction models of treatment outcome in MDD

The endeavor of finding indicators of treatment efficacy in MDD has led to a remarkable amount of publications from different psychiatric subfields. A large subset of these have looked at associations of preselected psychological and biological factors with treatment outcome. The main aim hereby was the identification of new (bio)markers using classical statistical approaches, such as regression models with null hypothesis significance testing based on p-values of the investigated predictors. The results from these association studies have been summarized in several systematic reviews and meta-analyses, often focusing on selected data modalities (but see [51, 52]), such as sociodemographic and clinical measures [53], cognitive functioning [54], or blood biomarkers [55]. Table 2 provides a list of these publications grouped by data modality and by their ease of access in clinical practice. Overall, the most consistently identified and most predictive factors were derived from sociodemographic and clinical characteristics [19]. Information on a patient’s social support, their baseline symptom severity, psychiatric comorbidities (e.g., anxiety disorders), or chronicity of the disorder, for instance, have repeatedly been associated with MDD treatment outcome [51–53]. However, an important shortcoming of these results is that none of the identified measures has been proven informative enough to sufficiently predict treatment outcome on their own.

This issue has led to a “new generation” of studies which aim at creating prediction models based on a multitude of variables. These models use machine learning (ML) methods, mainly supervised learning with classification algorithms such as regularized logistic regression or tree-based methods [56], to combine the effects of many variables and to increase predictive accuracy. Hence, they do not necessarily focus on the identification of new predictors of treatment outcome but rather try to find the best combination of variables to maximize their predictive power. A clear and comprehensive review on ML models and their value for predicting treatment outcome in psychiatry was recently published [57], as well as a systematic review and meta-analysis of these approaches in MDD specifically [58]. Crucially, the development of such models needs to include some kind of validation in order to assure that predictions are not specific to the data they were created from but also generalize to new data. Validation is often performed by dividing the initial data set into subsamples (e.g., training sample and validation sample) or by testing the model’s performance on a completely independent sample [59]. Furthermore, sufficiently large data sets in terms of sample size are required

Table 2 Different measurement techniques used in psychiatric research and corresponding examples of derived factors associated with antidepressant treatment effects

Measurement technique	Requirements	Example indicators of antidepressant treatment outcome
Easily accessible and usable		
Questionnaires and clinician-based ratings or interviews	Manuals (Clinical training)	Social demographics [53, 119, 120] Symptom profiles [53, 119, 120] Comorbidities [121, 122] Personality traits [123] Exposure to environmental risk factors, e.g., childhood abuse [124–126]
Tests and tasks	Manuals Technical devices for digital implementations	Cognitive functioning [54] Emotional processing [127, 128]
Technically feasible but additional efforts and expenses needed		
Blood draw or saliva sampling for established parameters	Medical training and equipment Laboratory capacities	Immune parameters, e.g., cytokines [129–131] Metabolites [132] Pharmacogenomic testing [133–135]
Dynamic function tests	Medical training and equipment Laboratory capacities	HPA-axis regulation [136, 137]
Technically feasible but high complexity and expenses		
Genotyping pipelines (based on blood draw or saliva sampling or other biospecimen)	Medical training and equipment Laboratory capacities Computational expertise and resources	Candidate genes without established testing [138, 139] Genome-wide associations [140–145] Polygenic risk scores [146] Epigenetic, transcriptomic and metabolomic markers [147–151]
Technical recording devices	Special equipment Technical training and expertise	Neuroimaging [152–155] Electroencephalographic markers [156, 157] Peripheral physiological markers [158]

Measurements are grouped by their accessibility and usability for routine clinical practice and licensed physicians. Note that this table is neither exhaustive nor based on a systematic literature search but meant to show exemplary indicators and their translational value

to guarantee robustness and generalizability of the predictions. The majority of predictive ML models of MDD treatment outcome have thus been created on data from large patient cohorts coming either from clinical trials (such as STAR*D [60, 61], Genome-based Therapeutic Drugs for Depression [62, 63], or Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care in Depression [64]), or from observational studies (such as the Munich Antidepressant Response Signature project [32] or the Netherlands Study of Depression and Anxiety (NESDA) [65, 66]). Since clinical trials usually compare different treatment arms (or treatment against placebo), the resulting predictions are likely to be treatment-specific and may not be readily applied to other treatments [60, 62]. Observational studies, on the other hand, follow a more naturalistic approach by observing patients who are treated based on routine clinical decisions, which might lead to more heterogeneity in the data [67, 68]. In general, prediction models of MDD treatment outcome based on sample sizes of at least several hundred patients (e.g., [60–63]) can predict treatment outcome (most often response vs. non-response or remission vs. non-remission) with moderate to good accuracies of

65%–75% [58]. This means that up to three quarters of ‘true’ responders/remitters are recognized as such by these prediction models. Most models that have been published so far have confirmed that the most reliable predictors of MDD treatment outcome come from established clinical and sociodemographic factors that had already been identified in earlier studies, such as initial symptom severity (e.g., [32, 36, 60, 62]), number and duration of depressive episodes (e.g., [32, 60]), personality traits (e.g., [32, 66]), as well as employment status and education (e.g., [61, 66]). However, only few studies exist that have assessed the additional value of other data modalities by comparing the performance of a multimodal model to a model using sociodemographic or clinical variables only. We here provide two examples of studies that have followed this approach using large sample sizes (at least several hundred samples) and ML methods. Iniesta et al. [63] showed that a prediction model combining demographic and clinical variables (e.g., depressive symptom scores, medication status, and stressful life events) with over 500,000 genetic markers (single nucleotide polymorphisms and copy number variants) led to slightly more accurate predictions (area under the receiver operating characteristic curve (AUC)

of 0.77) than a model trained on the non-genetic variables only (AUC of 0.74; [62]). Similarly, Dinga et al. [66] compared a prediction model combining clinical and biological data (primarily somatic health measures, inflammatory and metabolic markers) to models including only one of the available predictor domains. Across all comparisons, the full model containing all variables performed better than the alternative models. The largest differences occurred when the alternative model was based on biological measures only, the smallest differences when it was based on depressive symptom severity scores (differences in AUC of 0.01–0.05). These results suggest that even though adding biological markers to prediction models can lead to increases in performance, their additional value on top of clinical data still remains small.

Clinical decision support systems in psychiatry

A suitable instrument to transfer predictive models from research into clinical practice is a Clinical Decision Support System (CDSS). CDSSs are any kinds of computer systems that work with clinical data or knowledge and are set up to assist healthcare professionals in decision processes [69]. These decisions can refer to both diagnosing a patient and selecting the best treatment [70]. Concretely, a patient's characteristics enter a CDSS to be evaluated based on implemented clinical knowledge in order to return recommendations to the clinicians [71]. Hence, these systems can improve clinical processes and help healthcare professionals benefit from scientific findings [72].

CDSSs have been used successfully in many medical disciplines (for a review, see [73]), but use in psychiatry or mental health is lagging behind. However, some systems have been developed for the diagnoses of mental disorders, e.g., for attention deficit hyperactivity disorder [74], MDD and anxiety disorders [75], subtypes of schizophrenia [76], or a broader range of disorders [77]. Other systems were designed more specifically and can also be of value for MDD, such as the NetDSS [78], a web-based CDSS with various functions, from patient registry to clinical outcome monitoring. An elegant tool for physicians and patients was set up by Henshall et al. [79]. They developed a recommendation system and tested it on a focus group comprising physicians, caregivers, and patients with several mental disorders, including MDD. By entering basic sociodemographic and clinical variables as well as by setting preferences for potential side effects, the software returned a graphical illustration of recommended interventions and their corresponding probabilities of effectiveness. A benefit of such a tool is that it uses individual data to tailor a treatment to each

patient. Similarly, a few commercial tools have been developed lately, promoting improvements of treatment efficacy for mental disorders using individual patient data and predictive models [80–82].

Ultimately, such predictive systems can enhance personalized treatment, e.g., by indicating from the beginning which medication has the highest probability to lead to a beneficial response. Moreover, these tools can save physicians time and increase preciseness of clinical judgements [83, 84].

Current challenges and unmet needs

With the increasing interest in precision psychiatry and outcome prognosis, many efforts have been invested in this field of research. Nonetheless, the core problem in translational psychiatry remains: translations of research findings into daily clinical work, in such a way that patients and clinicians could directly benefit from them, are practically non-existent. Due to the lack of validated tests as guidance for personalized medication, treatment administration still has to rely on generic guidelines and physicians' personal judgements. The potential solution appears to be twofold: first, robust (bio)markers of treatment efficacy need to be identified and built into prognostic models. Subsequently, if models are proven useful, the second step will be their translation into new tools for clinicians. The main issues and current challenges in this translational process as well as potential solution approaches are outlined below. Additionally, they are illustrated in Fig. 1.

Challenges in concepts and definitions

Up to 16,400 potential symptom combinations can lead to a diagnosis of MDD [85], which might essentially be a conglomerate of many different pathophysiologies [86]. Moreover, MDD shows a high degree of comorbidity with other mental disorders, both cross-sectionally [87–89] and over time [90]. Longitudinal studies, especially using registry data [91], have shown large variability of diagnoses across lifetime which is why a cross-sectional focus on MDD diagnosis might miss relevant longitudinal information that discriminates among disorder subtypes. Hence, transdiagnostic and longitudinal approaches (e.g., assessing lifetime disorders in diagnostic interviews) should be considered in clinical studies.

A second challenge is posed by the measurements and definitions of antidepressant outcome (see Table 1). Unlike other medical disciplines, which provide objective biological measures of disease severity or treatment success, psychiatry defines clinical outcomes on subjective ratings (self-reported or clinician-rated). However, some of the

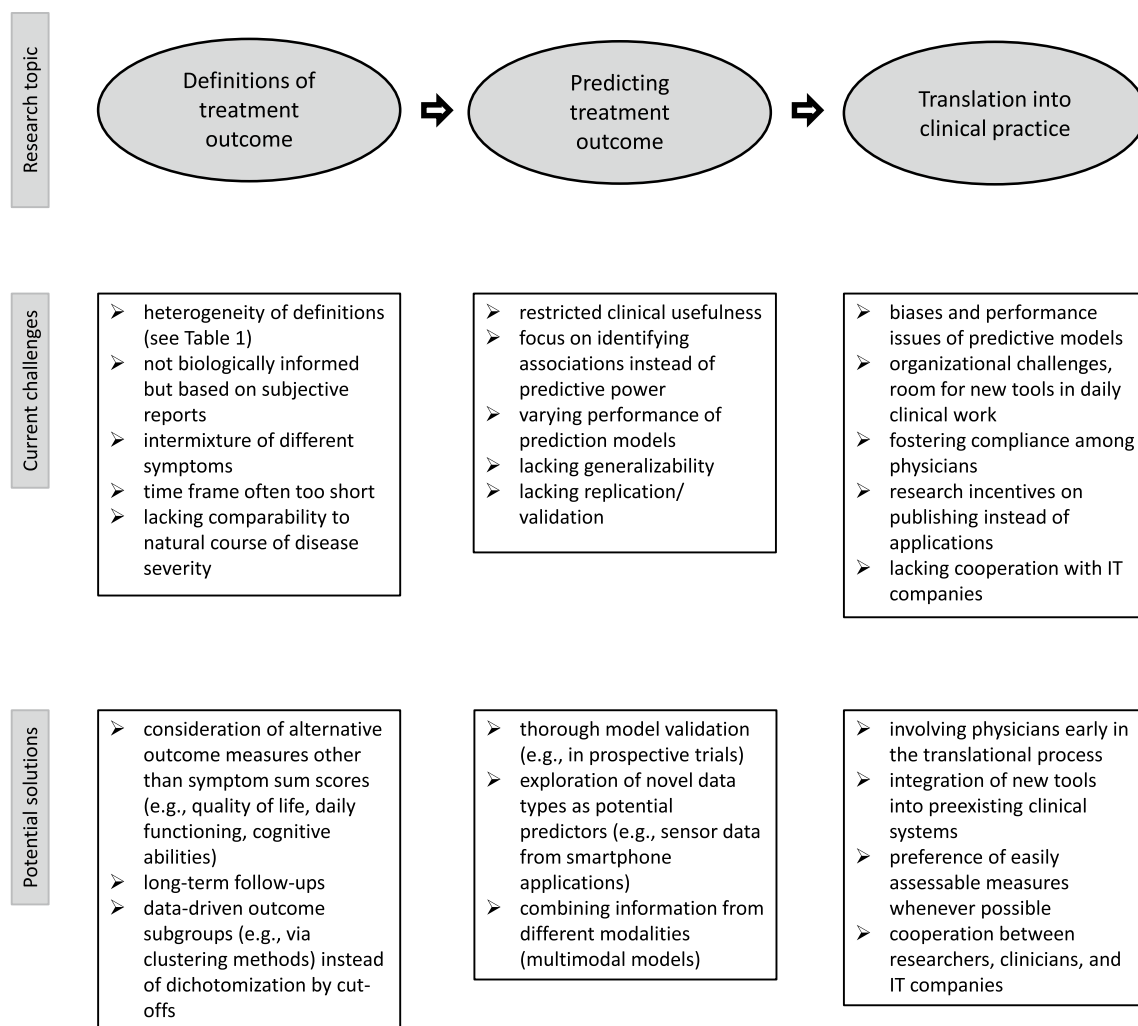


Fig. 1 Current challenges with respect to different stages of research on treatment outcome in MDD patients and its translation into clinical practice

most common ratings were shown to lack reliability [27, 92, 93] and to be incongruent among themselves, meaning that they do not measure exactly the same construct and are thus not fully comparable [28]. These issues limit the validity of findings and the generalizability from one outcome scale to others. Moreover, ratings of depressive symptom severity, such as the HDRS, the QIDS, or the BDI, evaluate many different symptoms and aspects of MDD, all influencing the respective sum score. It is possible for patients to show a 50% reduction of the sum score and be classified as responders, even when none of the core symptoms of MDD (depressed mood or reduced interest/pleasure in activities) have improved. Furthermore, patients with the same overall severity score can show very different symptom profiles, and have thus very different subjective experiences of their disorder. This important information gets lost when sum score data are used [94]. Explicitly differentiating between symptoms instead of using sum scores could help to identify

indicators of specific symptoms and could thus lead toward more targeted treatments [95].

Moreover, antidepressant outcome is often defined as (partial) response or remission (see Table 1). Both terms represent artificially dichotomized variables, created based on more or less arbitrary cut-off values on a continuous scale, that is, the respective sum score (for remission) or the difference in sum scores (for response) on a symptom scale. Dichotomizing continuous variables always brings certain risks and comes with loss of information [96]. Consider two patients with very similar symptom scores during the course of treatment, e.g., symptom reductions of 55% and 45%, respectively. According to the common definition of treatment response, the first patient would be classified as a ‘responder’ whereas the second patient would be treated as a ‘non-responder’. In fact, the second patient would be categorized together with patients who do not show any symptom reduction at all. Classifying patients in a data-driven manner,

e.g., using clustering techniques to create more homogenous outcome classes, might be a promising alternative that has already been implemented in several studies [32–35]. Still, the resulting outcome groups strongly depended on the selected variables and the chosen clustering method. Hence, the number of identified groups varied, e.g., from five [33] to seven [32, 34] up to nine [35]. These discrepancies challenge their clinical usefulness as the obtained classes are likely not generalizable to most other settings. Nevertheless, especially if more than one type of outcome measure is available, clustering methods might be a good way to combine information and identify subgroups.

Another issue with common measurements of treatment efficacy is the time frame. Patients in clinical trials are often measured over a few weeks only. Especially in disorders such as MDD, which can appear recurrently and show a risk of chronification [97], it is important to follow up on patients after a longer period of time. This could help differentiate between temporary improvements and long-term recovery. In the NESDA sample, 22% of initially remitted patients developed a recurrent episode within the following 2 years [98]. Identifying these at-risk patients early on might help to prevent subsequent episodes by scheduling regular checkups and implementing prevention strategies [99].

Even in the absence of reliable (biological) alternatives, sum scores on symptom questionnaires alone do not seem to be the most specific and clinically meaningful measures [95, 100]. In a recent online survey, MDD patients, informal caregivers, and healthcare professionals were asked to indicate outcome domains that matter most in their opinion. They identified not only depressive symptoms but also domains of functioning, healthcare organization, and social representation, many of which are not measured in most clinical studies, let alone included in depression rating scales [44, 101], highlighting the importance of including patient centered outcomes. Another research team explicitly differentiated between opinions from doctors and patients [102]. Their survey revealed that physicians mainly considered alleviation of depressive symptoms to be most important for relief and cure from MDD whereas patients rather focused on rehabilitation of positive affect. These results suggest that definitions and measures of treatment outcome should go beyond plain ratings of symptom changes and need to be broadened and potentially lengthened [42]. Relevant assessment instruments for many different domains of MDD characterization, including neurocognition, functioning and quality of life, as well as their suitability for routine clinical use have recently been reviewed [19] and should be considered when measuring treatment outcome in future studies.

Finally, novel objective measures that do not rely on subjective self- or external reports, such as behavioral and functional data generated by smartphones, wearables or other digital devices, could be of further value [103]. As long as

no direct biological measure of treatment outcome exists, personal data collected from mobile devices, i.e., ‘digital phenotyping’, might become a promising alternative [104]. Ecological momentary assessments, actimetry, speech characteristics, or movement patterns, for instance, can be continuously and mainly passively collected in large amounts and in high temporal resolution. Sensor data and other information from wearable devices like smartphones have already been successfully applied in psychiatric research, especially in combination with ML and deep learning [103]. Future studies will need to prove if they can contribute to a deeper and broader characterization of treatment outcome and MDD.

Challenges for prediction models

Except for a few psychometric and sociodemographic factors, there are still no robust or well replicated predictors of treatment outcome. Apart from a few promising pharmacogenetic tests [81, 105], no biological measures qualify as stable biomarkers nor are they used in clinical practice. Associations between specific measurements and treatment outcome are often of limited prognostic value as statistically significant associations do not guarantee accurate and robust predictions. Therefore, the focus has started to shift from testing associations to improving predictions in order to forecast what is most beneficial for an individual patient and to personalize clinical decision-making [106].

Predictive ML models tackle this issue as they are built to be as accurate and robust as possible. The robustness of a model should be assessed by validating it on an independent data set [57], ideally by testing its performance and safety on new patients in a prospective clinical study. Nonetheless, several prediction models were not validated on external data sets at all (e.g., [32, 64, 65]). Others were less predictive when they were applied to other classes of antidepressants, suggesting that the identified predictors of treatment outcome might be agent-specific [60, 62]. In addition, the main target variables in studies using ML were response and remission in their binary form [58], the downsides of which have already been discussed. Furthermore, psychiatric data often face the problem of high dimensionality while samples sizes remain relatively small [107]. This is often referred to as the ‘curse of dimensionality’: the more variables a data set contains, the more the sample size needs to increase (per variable) to allow reliable results [59]. Otherwise, resulting prediction models are likely to be biased and therefore need to be carefully validated on independent data to ensure their reliability. Moreover, prediction models based on biological data often only show restricted translatability into clinical practice as they require precisely preprocessed data from time-consuming and expensive measurements. A prerequisite for a successful translation of a predictive model into

clinical practice is that it consists of parameters that can be routinely accessed by a licensed physician without producing a lot of extra costs. Psychological and clinical features as well as sociodemographic information can be evaluated easily by any trained clinician or via self-ratings. On the other hand, as indicated in Table 2, many biological measures, i.e., potential biomarkers, are comparatively expensive or hard to assess for physicians in common clinical settings. This is especially the case for neuroimaging, omics data, and endocrinological markers derived from a challenge test, for instance. Such parameters should only be preferred over less costly data modalities, e.g., questionnaire data, if their predictive performance is notably higher and thus justifies the additional expenses. Making use of other objective measures, such as data collected from smartphones and other wearable devices, might become a promising alternative [103]. Their collection would be economical and profitable for researchers as well as less time-consuming and free of stress for patients.

In summary, well-performing and externally validated ML models are promising tools for future psychiatric practice [59], including the prognosis of treatment outcome in MDD.

Challenges for CDSSs

In order to translate predictive models into digital tools for everyday clinical use, CDSSs could be of help. Iniesta et al. [108] sketched a concise outline of the workflow for designing and choosing predictive models and, crucially, explained how to bring them into CDSSs. Still, as appealing as the idea of such publicly used tools might sound, they have not yet become prevalent in healthcare institutions.

The main challenge in MDD outcome prediction seems to be the lack of powerful models and established predictive patient characteristics. As outlined above, predictive models are still not robust and generalizable enough to guide daily clinical decisions. Only if additional value coming from a predictive model is proven, will an implementation into a CDSSs lead to a successful supporting device. Biases in such systems, for instance, were shown to lead to underestimations of their effectiveness [109], high non-compliance rates among users [73], and even to wrong diagnoses by physicians [110]. This is particularly concerning given that working with a CDSS might influence clinicians in their decisions later on even when they are not explicitly using the system anymore [111].

Furthermore, before CDSSs can be fully implemented into clinical workflows, substantial ethical challenges need to be considered. Apart from data protection, which needs to be assured, questions regarding liability and responsibility for treatment decisions have to be addressed, especially when it comes to disagreement between physicians and

support systems. Also, human interactions, conversations and relations between patients and mental health professionals play an important role, not only in psychiatric care [112, 113]. Further necessary ethical considerations have been summarized by Chekroud et al. [57].

Due to these problems, a number of factors needed to sustainably establish CDSSs in clinical settings should be considered [73]: First, apart from having appealing visual designs and being user-friendly, the system should implement personalized, transparent, and reliable recommendations as well as comprehensive overviews for each patient. Second, physicians should keep the authority over treatment decisions and should still oversee algorithmic outputs [114]. They should be involved in the development of the system, receive training and not have to make adaptations in their daily working processes in order to use the application. Third, to circumvent organizational obstacles, CDSSs should be integrated into preexisting clinical computerized systems, such as electronic medical records or physician order entries [73].

Ultimately, however, the main incentive in research seems to remain the publication of novel findings, indeed funding for the translation of existing findings into applications and technical devices is often more difficult to obtain [115, 116]. Therefore, interdisciplinary work is needed, bringing together scientists, clinicians and, e.g., information technologists for successful development of CDSSs.

Conclusion

Tackling the medical treatment of MDD and increasing treatment efficacy have always been major challenges in psychiatric research. In this narrative review, we summarized current approaches to operationalize and predict treatment outcome in MDD. We highlighted findings from ML approaches and discussed their implementation into CDSSs. To date, numerous studies have investigated and discovered associations between biological and phenotypic patient characteristics and treatment outcome, producing growing evidence for potential underlying mechanisms. Large patient cohort data and ML methods have additionally produced predictive models with promising accuracies (e.g., [32, 36, 60, 62, 64, 65]). Nevertheless, psychiatry has made comparatively little progress in applying the acquired knowledge into daily clinical work and in personalizing decisions based on empirically derived patient characteristics.

The main issue of this lacking translation seems to be the absence of robust and generalizable predictors of treatment outcome, especially of biological and other objectively measurable markers. Further quantitative characterizations of patients might help to identify more robust predictors and could provide support in medical decisions, such as choosing

the most beneficial treatment for individual patients or subgroups of patients [117]. Once reliable indicators and prognostic models are established, the next challenge will be their implementation into clinical practice. Efficient systems with clear interpretation of results need to be introduced and made available for healthcare professionals. CDSSs can be useful tools to implement tests and predictive models to guarantee benefits for physicians and patients. To make this happen, research funding needs to put more emphasis on translational systems, i.e., the development of target-oriented and clinically useful applications. Cooperation with companies specialized in health information technologies might be of particular use for this endeavor. Finally, there needs to be a shift in psychiatry toward a data-driven stratification of patients as well as more precise, personalized treatments based on individual patient data.

Acknowledgements We thank Dr. Jessica Keverne for proof-reading the article.

Author contributions NR wrote the initial draft of the manuscript. EBB and TMB critically contributed to the writing of the manuscript and its revisions. All authors contributed to and have approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This publication is funded by the Max Planck Institute of Psychiatry. NR is supported by the International Max Planck Research School of Translational Psychiatry (IMPRS-TP) and received funding from the Bavarian Ministry of Economic Affairs, Regional Development and Energy (BayMED, PBN_MED-1711–0003).

Declarations

Conflict of interest EBB is an editor of the journal *European Archives of Psychiatry and Clinical Neuroscience*. Otherwise, the authors have no financial or non-financial competing interests to declare that are relevant to the content of this article.

Employment All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. World Health Organization (2017) Depression and other common mental disorders: global health estimates. World Health Organization, Geneva
2. GBD (2016) Disease and Injury Incidence and Prevalence Collaborators (2017) Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 390:1211–1259. [https://doi.org/10.1016/S0140-6736\(17\)32154-2](https://doi.org/10.1016/S0140-6736(17)32154-2)
3. Thomas L, Kessler D, Campbell J et al (2013) Prevalence of treatment-resistant depression in primary care: cross-sectional data. *Br J Gen Pract* 63:852–858. <https://doi.org/10.3399/bjgp13X675430>
4. Khan A, Fahl Mar K, Faucett J et al (2017) Has the rising placebo response impacted antidepressant clinical trial outcome? Data from the US food and drug administration 1987–2013. *World Psychiatry* 16:181–192. <https://doi.org/10.1002/wps.20421>
5. Trivedi MH, Rush AJ, Wisniewski SR et al (2006) Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 163:28–40. <https://doi.org/10.1176/appi.ajp.163.1.28>
6. Bingham KS, Rothschild AJ, Mulsant BH, et al (2017) The Association of Baseline Suicidality With Treatment Outcome in Psychotic Depression. *J Clin Psychiatry* 78:1149–1154. <https://doi.org/10.4088/JCP.14m09658>
7. Greenberg PE, Fournier AA, Sisitsky T et al (2015) The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry* 76:155–162. <https://doi.org/10.4088/JCP.14m09298>
8. Cohen ZD, DeRubeis RJ (2018) Treatment selection in depression. *Annu Rev Clin Psychol* 14:209–236. <https://doi.org/10.1146/annurev-clinpsy-050817-084746>
9. Friston KJ, Redish AD, Gordon JA (2017) Computational nosology and precision psychiatry. *Comput psychiatry (Cambridge, Mass)* 1:2–23. https://doi.org/10.1162/CPSY_a_00001
10. DeRubeis RJ, Siegle GJ, Hollon SD (2008) Cognitive therapy versus medication for depression: treatment outcomes and neural mechanisms. *Nat Rev Neurosci* 9:788–796. <https://doi.org/10.1038/nrn2345>
11. Collins FS, Varmus H (2015) A new initiative on precision medicine. *N Engl J Med* 372:793–795. <https://doi.org/10.1056/NEJMp1002530>
12. National Research Council (2011) Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. The National Academies Press, Washington, D.C.
13. Kraus C, Kadriu B, Lanzenberger R, et al (2019) Prognosis and improved outcomes in major depression: a review. *Transl Psychiatry* 9:. <https://doi.org/10.1038/s41398-019-0460-3>
14. American Psychiatric Association (2000) Practice guideline for the treatment of patients with major depressive disorder (revision). *Am J Psychiatry* 157:1–45
15. DGPPN, BÄK, KBV, et al (2015) S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression–Langfassung, 1. Auflage. Version 5
16. Hong RH, Murphy JK, Michalak EE et al (2021) Implementing measurement-based care for depression: practical solutions for psychiatrists and primary care physicians. *Neuropsychiatr Dis Treat* 17:79–90. <https://doi.org/10.2147/NDT.S283731>
17. Härter M, Watzke B, Daubmann A et al (2018) Guideline-based stepped and collaborative care for patients with depression in a cluster-randomised trial. *Sci Rep* 8:1–9. <https://doi.org/10.1038/s41598-018-27470-6>

18. McHugh RK, Whitton SW, Peckham AD et al (2013) Patient preference for psychological vs. pharmacological treatment of psychiatric disorders: a meta-analytic review. *J Clin Psychiatry* 74:595–602. <https://doi.org/10.4088/JCP.12r07757.Patient>
19. Maj M, Stein DJ, Parker G et al (2020) The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry* 19:269–293. <https://doi.org/10.1002/wps.20771>
20. Malhi GS, Das P, Mannie Z, Irwin L (2019) Treatment-resistant depression: problematic illness or a problem in our approach? *Br J Psychiatry* 214:1–3. <https://doi.org/10.1192/bjp.2018.246>
21. Hamilton M (1960) A rating scale for depression. *J Neurol Neurosurg Psychiatry* 23:56–62
22. Montgomery SA, Åsberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* 134:382–389. <https://doi.org/10.1192/bjp.134.4.382>
23. Rush AJ, Trivedi MH, Ibrahim HM et al (2003) The 16-item Quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 54:573–583. [https://doi.org/10.1016/S0006-3223\(02\)01866-8](https://doi.org/10.1016/S0006-3223(02)01866-8)
24. Beck AT, Ward C, Mendelson M et al (1961) Beck depression inventory (BDI). *Arch Gen Psychiatry* 4:561–571. <https://doi.org/10.1093/ndt/gfr086>
25. Williams JBW, Kobak KA (2008) Development and reliability of a structured interview guide for the Montgomery-Åsberg Depression Rating Scale (SIGMA). *Br J Psychiatry* 192:52–58. <https://doi.org/10.1192/bjp.bp.106.032532>
26. Hooijer C, Zitman FG, Griez E et al (1991) The Hamilton Depression Rating Scale (HDRS): changes in scores as a function of training and version used. *J Affect Disord* 22:21–29. [https://doi.org/10.1016/0165-0327\(91\)90079-8](https://doi.org/10.1016/0165-0327(91)90079-8)
27. Uher R, Farmer A, Maier W et al (2008) Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med* 38:289–300. <https://doi.org/10.1017/S00332917001730>
28. Uher R, Perlis RH, Placentino A et al (2012) Self-report and clinician-rated measures of depression severity: Can one replace the other? *Depress Anxiety* 29:1043–1049. <https://doi.org/10.1002/da.21993>
29. Rush AJ, Kraemer HC, Sackeim HA et al (2006) Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology* 31:1841–1853. <https://doi.org/10.1038/sj.npp.1301131>
30. Paykel ES, Ramana R, Cooper Z et al (1995) Residual symptoms after partial remission: an important outcome in depression. *Psychol Med* 25:1171–1180. <https://doi.org/10.1017/S003329170033146>
31. Thase ME (2003) Achieving remission and managing relapse in depression. *J Clin Psychiatry* 64(Suppl 1):3–7
32. Paul R, Andlauer TFM, Czamara D, et al (2019) Treatment response classes in major depressive disorder identified by model-based clustering and validated by clinical prediction models. *Transl Psychiatry* 9:. <https://doi.org/10.1038/s41398-019-0524-4>
33. Kelley ME, Dunlop BW, Nemeroff CB et al (2018) Response rate profiles for major depressive disorder: characterizing early response and longitudinal nonresponse. *Depress Anxiety* 35:992–1000. <https://doi.org/10.1002/da.22832>
34. Hartmann A, von Wietersheim J, Weiss H, Zeeck A (2018) Patterns of symptom change in major depression: classification and clustering of long term courses. *Psychiatry Res* 267:480–489. <https://doi.org/10.1016/j.psychres.2018.03.086>
35. Uher R, Mors O, Rietschel M et al (2011) Early and delayed onset of response to antidepressants in individual trajectories of change during treatment of major depression: a secondary analysis of data from the genome-based therapeutic drugs for depression (GENDEP) study. *J Clin Psychiatry* 72:1478–1484. <https://doi.org/10.4088/JCP.10m06419>
36. Athreya AP, Brückl T, Binder EB et al (2021) Prediction of short-term antidepressant response using probabilistic graphical models with replication across multiple drugs and treatment settings. *Neuropsychopharmacology*. <https://doi.org/10.1038/s41386-020-00943-x>
37. McIntyre RS, Filteau MJ, Martin L et al (2014) Treatment-resistant depression: definitions, review of the evidence, and algorithmic approach. *J Affect Disord* 156:1–7. <https://doi.org/10.1016/j.jad.2013.10.043>
38. Berlim MT, Turecki G (2007) Definition, assessment, and staging of treatment-resistant refractory major depression: a review of current concepts and methods. *Can J Psychiatry* 52:46–54
39. Souery D, Amsterdam J, De Montigny C et al (1999) Treatment resistant depression: methodological overview and operational criteria. *Eur Neuropsychopharmacol* 9:83–91. [https://doi.org/10.1016/S0924-977X\(98\)00004-2](https://doi.org/10.1016/S0924-977X(98)00004-2)
40. Anderson IM (2018) We all know what we mean by treatment-resistant depression—Don't we? *Br J Psychiatry* 212:259–261. <https://doi.org/10.1192/bjp.2018.56>
41. Conway CR, George MS, Sackeim HA (2017) Toward an evidence-based, operational definition of treatment-resistant depression: When enough is enough. *JAMA Psychiat* 74:9–10. <https://doi.org/10.1001/jamapsychiatry.2016.2586>
42. Slofstra C, Booij SH, Rogier Hoenders HJ, Castelein S (2019) Redefining therapeutic outcomes of depression treatment. *J Pers Res* 5:115–122. <https://doi.org/10.17505/jpor.2019.10>
43. Ishak WW, Greenberg JM, Balayan K et al (2011) Quality of life: the ultimate outcome measure of interventions in major depressive disorder. *Harv Rev Psychiatry* 19:229–239. <https://doi.org/10.3109/10673229.2011.614099>
44. McKnight PE, Kashdan TB (2009) The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. *Clin Psychol Rev* 29:243–259. <https://doi.org/10.1016/j.cpr.2009.01.005>
45. Lee RSC, Hermens DF, Porter MA, Redoblado-Hodge MA (2012) A meta-analysis of cognitive deficits in first-episode Major Depressive Disorder. *J Affect Disord* 140:113–124. <https://doi.org/10.1016/j.jad.2011.10.023>
46. Zimmerman M, McGlinchey JB, Posternak MA et al (2006) How should remission from depression be defined? The depressed patient's perspective. *Am J Psychiatry* 163:148–150
47. McIntyre RS, Cha DS, Soczynska JK et al (2013) Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. *Depress Anxiety* 30:515–527. <https://doi.org/10.1002/da.22063>
48. American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC
49. Endicott J, Nee J, Harrison W, Blumenthal R (1993) Quality of life enjoyment and satisfaction questionnaire: a new measure. *Psychopharmacol Bull* 29:321–326
50. Üstün TB, Chatterji S, Kostanjsek N et al (2010) Developing the World Health Organization disability assessment schedule 2.0. *Bull World Health Organ* 88:815–823. <https://doi.org/10.2471/BLT.09.067231>
51. Bennabi D, Aouizerate B, El-Hage W et al (2015) Risk factors for treatment resistance in unipolar depression: a systematic review. *J Affect Disord* 171:137–141. <https://doi.org/10.1016/j.jad.2014.09.020>

52. Perlman K, Benrimoh D, Israel S et al (2019) A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. *J Affect Disord* 243:503–515. <https://doi.org/10.1016/j.jad.2018.09.067>
53. De Carlo V, Calati R, Serretti A (2016) Socio-demographic and clinical predictors of non-response/non-remission in treatment resistant depressed patients: a systematic review. *Psychiatry Res* 240:421–430. <https://doi.org/10.1016/j.psychres.2016.04.034>
54. Pimontel MA, Rindskopf D, Rutherford BR et al (2016) A meta-analysis of executive dysfunction and antidepressant treatment response in late-life depression. *Am J Geriatr Psychiatry* 24:31–41. <https://doi.org/10.1161/CIRCULATIONAHA.114.010270>
55. Mora C, Zonca V, Riva MA, Cattaneo A (2018) Blood biomarkers and treatment response in major depression. *Expert Rev Mol Diagn* 18:513–529. <https://doi.org/10.1080/14737159.2018.1470927>
56. Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edi. Springer, New York, NY
57. Chekroud AM, Bondar J, Delgado J et al (2021) The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry* 20:154–170. <https://doi.org/10.1002/wps.20882>
58. Lee Y, Ragguett RM, Mansur RB et al (2018) Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *J Affect Disord* 241:519–532. <https://doi.org/10.1016/j.jad.2018.08.073>
59. Dwyer DB, Falkai P, Koutsouleris N (2018) Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol* 14:91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
60. Chekroud AM, Zotti RJ, Shehzad Z et al (2016) Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry* 3:243–250. [https://doi.org/10.1016/S2215-0366\(15\)00471-X](https://doi.org/10.1016/S2215-0366(15)00471-X)
61. Jain FA, Hunter AM, Brooks JO, Leuchter AF (2013) Predictive socioeconomic and clinical profiles of antidepressant response and remission. *Depress Anxiety* 30:624–630. <https://doi.org/10.1002/da.22045>
62. Iniesta R, Malki K, Maier W et al (2016) Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *J Psychiatr Res* 78:94–102. <https://doi.org/10.1016/j.jpsyc.2016.03.016>
63. Iniesta R, Hodgson K, Stahl D et al (2018) Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables. *Sci Rep* 8:1–9. <https://doi.org/10.1038/s41598-018-23584-z>
64. Wu W, Zhang Y, Jiang J et al (2020) An electroencephalographic signature predicts antidepressant response in major depression. *Nat Biotechnol* 38:439–447. <https://doi.org/10.1038/s41587-019-0397-3>
65. Frässle S, Marquand AF, Schmaal L et al (2020) Predicting individual clinical trajectories of depression with generative embedding. *NeuroImage Clin* 26:102213. <https://doi.org/10.1016/j.nicl.2020.102213>
66. Dinga R, Marquand AF, Veltman DJ et al (2018) Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Transl Psychiatry* 8:241. <https://doi.org/10.1038/s41398-018-0289-1>
67. Ross JS (2014) Randomized clinical trials and observational studies are more often alike than unlike. *JAMA Intern Med* 174:1557. <https://doi.org/10.1001/jamainternmed.2014.3366>
68. Webb CA, Cohen ZD, Beard C et al (2020) Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: a comparison of machine learning approaches. *J Consult Clin Psychol* 88:25–38. <https://doi.org/10.1037/ccp0000451>
69. Kemppinen J, Korpela J, Elfvengren K, et al (2014) Decision Support in Evaluating the Impacts of Mental Disorders on Work Ability. 2014 47th Hawaii Int Conf Syst Sci 2958–2966. <https://doi.org/10.1109/HICSS.2014.368>
70. Musen MA, Middleton B, Greenes RA (2014) Clinical decision-support systems. In: Shortliffe E, Cimino J (eds) *Biomedical informatics*. Springer, London, pp 643–674
71. Sim I, Gorman P, Greenes RA et al (2001) Clinical decision support systems for the practice of evidence-based medicine. *J Am Med Assoc* 285:527–534
72. Bright TJ, Wong A, Dhurjati R et al (2012) Effect of clinical decision-support systems: a systematic review. *Ann Intern Med* 157:29–43. <https://doi.org/10.7326/0003-4819-157-1-201207030-00450>
73. Kilsdonk E, Peute LW, Jaspers MWM (2017) Factors influencing implementation success of guideline-based clinical decision support systems: a systematic review and gaps analysis. *Int J Med Inform* 98:56–64. <https://doi.org/10.1016/j.ijmedinf.2016.12.001>
74. Kemppinen J, Korpela J, Elfvengren K, et al (2013) A Clinical Decision Support System for adult ADHD diagnostics process. *Proc Annu Hawaii Int Conf Syst Sci* 2616–2625. <https://doi.org/10.1109/HICSS.2013.30>
75. Suhasini A, Palanivel S, Ramalingam V (2011) Multimodel decision support system for psychiatry problem. *Expert Syst Appl* 38:4990–4997. <https://doi.org/10.1016/j.eswa.2010.09.152>
76. Razzouk D, Mari JJ, Shirakawa I, et al (2006) Decision support system for the diagnosis of schizophrenia disorders. *Brazilian J Med Biol Res = Rev Bras Pesqui medicas e Biol* 39:119–128. S0100-879X2006000100014
77. Bergman LG, Fors UGH (2008) Decision support in psychiatry—a comparison between the diagnostic outcomes using a computerized decision support system versus manual diagnosis. *BMC Med Inform Decis Mak* 8:9. <https://doi.org/10.1186/1472-6947-8-9>
78. Fortney JC, Pyne JM, Steven CA et al (2010) A web-based clinical decision support system for depression care management. *Am J Manag Care* 16:849–954. <https://doi.org/10.1016/j.atheroscclerosis.2009.05.009.Effect>
79. Henshall C, Marzano L, Smith K 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* 17:265. <https://doi.org/10.1186/s12888-017-1406-z>
80. aifred health (2020) Aifred Health. <https://aifredhealth.com/>
81. Assurex Health Inc. (2020) GeneSight. Changing lives through genetic insight. <https://genesight.com/>
82. Spring Care Inc. (2020) Spring Health. <https://www.springhealth.com/>
83. Dawes RM, Faust D, Meehl PE (1989) Clinical versus actuarial assessments. *Science* (80-) 243:1668–1674
84. Trivedi MH, Rush AJ, Crismon ML et al (2004) Clinical results for patients with major depressive disorder in the Texas medication algorithm project. *Arch Gen Psychiatry* 61:669–680. <https://doi.org/10.1001/archpsyc.61.7.669>
85. Fried EI, Nesse RM (2015) Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. *J Affect Disord* 172:96–102. <https://doi.org/10.1016/j.jad.2014.10.010>
86. Olbert CM, Gala GJ, Tupler LA (2014) Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. *J Abnorm Psychol* 123:452–462. <https://doi.org/10.1037/a0036068>

87. Hasin DS, Sarvet AL, Meyers JL et al (2018) Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiat* 75:336–346. <https://doi.org/10.1001/jamapsychiatry.2017.4602>
88. Lamers F, Van Oppen P, Comijs HC et al (2011) Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *J Clin Psychiatry* 72:341–348. <https://doi.org/10.4088/JCP.10m06176blu>
89. Jacobi F, Wittchen H-U, Höltling C et al (2004) Prevalence, comorbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). *Psychol Med* 34:597–611. <https://doi.org/10.1017/S00332917030001399>
90. Caspi A, Houts RM, Ambler A et al (2020) Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the dunedin birth cohort study. *JAMA Netw open* 3:e203221. <https://doi.org/10.1001/jamanetworkopen.2020.3221>
91. Plana-Ripoll O, Pedersen CB, Holtz Y et al (2019) Exploring comorbidity within mental disorders among a danish national population. *JAMA Psychiat* 76:259–270. <https://doi.org/10.1001/jamapsychiatry.2018.3658>
92. Trajković G, Starčević V, Latas M et al (2011) Reliability of the Hamilton rating scale for depression: a meta-analysis over a period of 49 years. *Psychiatry Res* 189:1–9. <https://doi.org/10.1016/j.psychres.2010.12.007>
93. Bagby RM, Ryder AG, Schuller DR, Marshall MB (2004) The Hamilton depression rating scale: Has the gold standard become a lead weight? *Am J Psychiatry* 161:2163–2177. <https://doi.org/10.1176/appi.ajp.161.12.2163>
94. McNeish D, Wolf MG (2020) Thinking twice about sum scores. *Behav Res Methods* 52:2287–2305. <https://doi.org/10.3758/s13428-020-01398-0>
95. Fried EI, Nesse RM (2015) Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Med* 13:1–11. <https://doi.org/10.1186/s12916-015-0325-4>
96. Altman DG, Royston P (2006) The cost of dichotomising continuous variables. *Br Med J* 332:1080. <https://doi.org/10.1136/bmj.332.7549.1080>
97. Hardeveld F, Spijker J, De Graaf R et al (2010) Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatr Scand* 122:184–191. <https://doi.org/10.1111/j.1600-0447.2009.01519.x>
98. Penninx BWJH, Nolen WA, Lamers F et al (2011) Two-year course of depressive and anxiety disorders: results from the Netherlands study of depression and anxiety (NESDA). *J Affect Disord* 133:76–85. <https://doi.org/10.1016/j.jad.2011.03.027>
99. Otte C, Gold SM, Penninx BW et al (2016) Major depressive disorder. *Nat Rev Dis Prim* 2:1–21. <https://doi.org/10.1038/nrdp.2016.65>
100. Fried EI (2017) The 52 symptoms of major depression: lack of content overlap among seven common depression scales. *J Affect Disord* 208:191–197. <https://doi.org/10.1016/j.jad.2016.10.019>
101. Chevance AM, Ravaud P, Tomlinson A et al (2020) Identifying outcomes for depression that matter to patients, informal caregivers and healthcare professionals: qualitative content analysis of a large international online survey. *Lancet Psychiatry* 7:692–702. [https://doi.org/10.1016/S2215-0366\(20\)30191-7](https://doi.org/10.1016/S2215-0366(20)30191-7)
102. Demyttenaere K, Donneau AF, Albert A et al (2015) What is important in being cured from depression? Discordance between physicians and patients (1). *J Affect Disord* 174:390–396. <https://doi.org/10.1016/j.jad.2014.12.004>
103. Durstewitz D, Koppe G, Meyer-Lindenberg A (2019) Deep neural networks in psychiatry. *Mol Psychiatry* 24:1583–1598. <https://doi.org/10.1038/s41380-019-0365-9>
104. Huckvale K, Venkatesh S, Christensen H (2019) Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *npj Digit Med* 2:. <https://doi.org/10.1038/s41746-019-0166-1>
105. Zeier Z, Carpenter LL, Kalin NH et al (2018) Clinical implementation of pharmacogenetic decision support tools for antidepressant drug prescribing. *Am J Psychiatry* 175:873–886. <https://doi.org/10.1176/appi.ajp.2018.17111282.Clinical>
106. Bzdok D, Varoquaux G, Steyerberg EW (2021) Prediction, not association, paves the road to precision medicine. *JAMA Psychiat* 78:127–128. <https://doi.org/10.1001/jamapsychiatry.2020.2549>
107. Rutledge RB, Chekroud AM, Huys QJ (2019) Machine learning and big data in psychiatry: toward clinical applications. *Curr Opin Neurobiol* 55:152–159. <https://doi.org/10.1016/j.conb.2019.02.006>
108. Iniesta R, Stahl D, McGuffin P (2016) Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol Med* 46:2455–2465. <https://doi.org/10.1017/S0033291716001367>
109. Tsai C-Y, Wang S-H, Hsu M-H, Li Y-C (2016) Do false positive alerts in naïve clinical decision support system lead to false adoption by physicians? A randomized controlled trial. *Comput Methods Programs Biomed* 132:83–91. <https://doi.org/10.1016/j.cmpb.2016.04.011>
110. Dreiseitl S, Binder M (2005) Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artif Intell Med* 33:25–30. <https://doi.org/10.1016/j.artmed.2004.07.007>
111. Browning M, Bilderbeck AC, Dias R et al (2021) The clinical effectiveness of using a predictive algorithm to guide antidepressant treatment in primary care (PREDicT): an open-label, randomised controlled trial. *Neuropsychopharmacology* 46:1307–1314. <https://doi.org/10.1038/s41386-021-00981-z>
112. Kelley JM, Kraft-Todd G, Schapira L, et al (2014) The influence of the patient-clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomized controlled trials. *PLoS One* 9:. <https://doi.org/10.1371/journal.pone.0094207>
113. Thompson L, McCabe R (2012) The effect of clinician-patient alliance and communication on treatment adherence in mental health care: A systematic review. *BMC Psychiatry* 12:. <https://doi.org/10.1186/1471-244X-12-87>
114. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 25:44–56. <https://doi.org/10.1038/s41591-018-0300-7>
115. Grimes DR, Bauch CT, Ioannidis JPA (2018) Modelling science trustworthiness under publish or perish pressure. *R Soc Open Sci* 5:. <https://doi.org/10.1098/rsos.171511>
116. Fanelli D (2010) Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS One* 5:. <https://doi.org/10.1371/journal.pone.0010271>
117. Akil H, Gordon J, Hen R et al (2018) Treatment resistant depression: A multi-scale, systems biology approach. *Neurosci Biobehav Rev* 84:272–288. <https://doi.org/10.1016/j.neubiorev.2017.08.019>
118. Houston JP, Gatz JL, Degenhardt EK, Jamal HH (2010) Symptoms predicting remission after divalproex augmentation with olanzapine in partially nonresponsive patients experiencing mixed bipolar i episode: a post-hoc analysis of a randomized controlled study. *BMC Res Notes* 3:1–6. <https://doi.org/10.1186/1756-0500-3-276>
119. DeRubeis RJ, Cohen ZD, Forand NR et al (2014) The personalized advantage index: translating research on prediction into individualized treatment recommendations. A demonstration *PLoS One* 9:1–8. <https://doi.org/10.1371/journal.pone.0083875>

120. Fournier JC, DeRubeis RJ, Hollon SD et al (2010) Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 303:175–177. <https://doi.org/10.1001/jama.2009.1943>
121. Souery D, Oswald P, Massat I et al (2007) Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. *J Clin Psychiatry* 68:1062–1070. <https://doi.org/10.4088/JCP.v68n0713>
122. Howland RH, Rush AJ, Wisniewski SR et al (2009) Concurrent anxiety and substance use disorders among outpatients with major depression: Clinical features and effect on treatment outcome. *Drug Alcohol Depend* 99:248–260. <https://doi.org/10.1016/j.drugalcdep.2008.08.010>
123. Takahashi M, Shirayama Y, Muneoka K et al (2013) Personality traits as risk factors for treatment-resistant depression. *PLoS One* 8:1–7. <https://doi.org/10.1371/journal.pone.0063756>
124. Nanni V, Uher R, Danese A (2012) Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *Am J Psychiatry* 169:141–151. <https://doi.org/10.1176/appi.ajp.2011.11020335>
125. Nelson J, Klumparendt A, Doeblen P, Ehring T (2017) Childhood maltreatment and characteristics of adult depression: meta-analysis. *Br J Psychiatry* 210:96–104. <https://doi.org/10.1192/bjp.bp.115.180752>
126. Williams LM, Debatista C, Duchemin AM et al (2016) Childhood trauma predicts antidepressant response in adults with major depression: data from the randomized international study to predict optimized treatment for depression. *Transl Psychiatry* 6:e799–e807. <https://doi.org/10.1038/tp.2016.61>
127. Godlewska BR, Browning M, Norbury R et al (2016) Early changes in emotional processing as a marker of clinical response to SSRI treatment in depression. *Transl Psychiatry* 6:e957–e967. <https://doi.org/10.1038/tp.2016.130>
128. Browning M, Kingslake J, Dourish CT et al (2019) Predicting treatment response to antidepressant medication using early changes in emotional processing. *Eur Neuropsychopharmacol* 29:66–75. <https://doi.org/10.1016/j.euroneuro.2018.11.1102>
129. Haroon E, Daguanno AW, Woolwine BJ et al (2018) Antidepressant treatment resistance is associated with increased inflammatory markers in patients with major depressive disorder. *Psychoneuroendocrinology* 95:43–49. <https://doi.org/10.1016/j.psyneuen.2018.05.026>
130. Liu JJ, Bin WY, Strawbridge R et al (2020) Peripheral cytokine levels and response to antidepressant treatment in depression: a systematic review and meta-analysis. *Mol Psychiatry* 25:339–350. <https://doi.org/10.1038/s41380-019-0474-5>
131. Zhou C, Zhong J, Zou B et al (2017) Meta-analyses of comparative efficacy of antidepressant medications on peripheral BDNF concentration in patients with depression. *PLoS One* 12:1–18. <https://doi.org/10.1371/journal.pone.0172270>
132. Kaddurah-Daouk R, Boyle SH, Matson W et al (2011) Pretreatment metabotype as a predictor of response to sertraline or placebo in depressed outpatients: a proof of concept. *Transl Psychiatry* 1:1–7. <https://doi.org/10.1038/tp.2011.22>
133. Altar CA, Carhart JM, Allen JD et al (2015) Clinical validity: combinatorial pharmacogenomics predicts antidepressant responses and healthcare utilizations better than single gene phenotypes. *Pharmacogenomics J* 15:443–451. <https://doi.org/10.1038/tpj.2014.85>
134. Bousman CA, Arandjelovic K, Mancuso SG et al (2019) Pharmacogenetic tests and depressive symptom remission: a meta-analysis of randomized controlled trials. *Pharmacogenomics* 20:37–47. <https://doi.org/10.2217/pgs-2018-0142>
135. Brown L, Vranjkovic O, Li J et al (2020) The clinical utility of combinatorial pharmacogenomic testing for patients with depression: a meta-analysis. *Pharmacogenomics* 21:559–569. <https://doi.org/10.2217/pgs-2019-0157>
136. Binder EB, Künzel HE, Nickel T et al (2009) HPA-axis regulation at in-patient admission is associated with antidepressant therapy outcome in male but not in female depressed patients. *Psychoneuroendocrinology* 34:99–109. <https://doi.org/10.1016/j.psyneuen.2008.08.018>
137. Fischer S, Macare C, Cleare AJ (2017) Hypothalamic-pituitary-adrenal (HPA) axis functioning as predictor of antidepressant response—meta-analysis. *Neurosci Biobehav Rev* 83:200–211. <https://doi.org/10.1016/j.neubiorev.2017.10.012>
138. Fabbri C, Corponi F, Souery D et al (2019) The genetics of treatment-resistant depression: a critical review and future perspectives. *Int J Neuropsychopharmacol* 22:93–104. <https://doi.org/10.1093/ijnpp/pyy024>
139. Uher R, Perroud N, Ng MYM et al (2010) Genome-wide pharmacogenetics of antidepressant response in the GENDEP Project. *Am J Psychiatry* 167:1–10. <https://doi.org/10.1176/appi.ajp.2009.09070932>
140. Adkins DE, Åberg K, McClay JL et al (2010) A genomewide association study of citalopram response in major depressive disorder—a psychometric approach. *Biol Psychiatry* 68:e25–e27. <https://doi.org/10.1016/j.biopsych.2010.05.018>
141. Biernacka JM, Sangkuhl K, Jenkins G et al (2015) The International SSRI Pharmacogenomics Consortium (ISPC): A genome-wide association study of antidepressant treatment response. *Transl Psychiatry* 5:1–9. <https://doi.org/10.1038/tp.2015.47>
142. Garriock HA, Kraft JB, Shyn SI et al (2010) A Genomewide Association Study of Citalopram Response in Major Depressive Disorder. *Biol Psychiatry* 67:133–138. <https://doi.org/10.1016/j.biopsych.2009.08.029>
143. GENDEP Investigators, MARS Investigators, STAR*D Investigators (2013) Common genetic variation and antidepressant efficacy in major depressive disorder: a meta-analysis of three genome-wide pharmacogenetic studies. *Am J Psychiatry* 170:207–217. <https://doi.org/10.1176/appi.ajp.2012.12020237>
144. Ising M, Lucae S, Binder EB et al (2009) A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch Gen Psychiatry* 66:966–975
145. Tansey KE, Guipponi M, Hu X et al (2013) Contribution of common genetic variants to antidepressant response. *Biol Psychiatry* 73:679–682. <https://doi.org/10.1016/j.biopsych.2012.10.030>
146. García-González J, Tansey KE, Hauser J et al (2017) Pharmacogenetics of antidepressant response: a polygenic approach. *Prog Neuro-Psychopharmacology Biol Psychiatry* 75:128–134. <https://doi.org/10.1016/j.pnpbp.2017.01.011>
147. Domschke K, Tidow N, Schwarte K et al (2014) Serotonin transporter gene hypomethylation predicts impaired antidepressant treatment response. *Int J Neuropsychopharmacol* 17:1167–1176. <https://doi.org/10.1017/S146114571400039X>
148. Lisoway AJ, Zai CC, Tiwari AK, Kennedy JL (2018) DNA methylation and clinical response to antidepressant medication in major depressive disorder: a review and recommendations. *Neurosci Lett* 669:14–23. <https://doi.org/10.1016/j.neulet.2016.12.071>
149. Belzeaux R, Lin R, Ju C et al (2018) Transcriptomic and epigenomic biomarkers of antidepressant response. *J Affect Disord* 233:36–44. <https://doi.org/10.1016/j.jad.2017.08.087>
150. Caspani G, Turecki G, Lam RW et al (2021) Metabolomic signatures associated with depression and predictors of antidepressant response in humans: A CAN-BIND-1 report. *Commun Biol* 4. <https://doi.org/10.1038/s42003-021-02421-6>
151. Alshehri T, Mook-Kanamori DO, Willems Van Dijk K et al (2021) Metabolomics dissection of depression heterogeneity and related cardiometabolic risk. *Psychol Med*. <https://doi.org/10.1017/S0033291721001471>

152. Fu CHY, Steiner H, Costafreda SG (2013) Predictive neural biomarkers of clinical response in depression: a meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiol Dis* 52:75–83. <https://doi.org/10.1016/j.nbd.2012.05.008>
153. Lener MS, Iosifescu DV (2015) In pursuit of neuroimaging biomarkers to guide treatment selection in major depressive disorder: a review of the literature. *Ann NY Acad Sci* 1344:50–65. <https://doi.org/10.1111/nyas.12759>
154. Enneking V, Leehr EJ, Dannlowski U, Redlich R (2019) Brain structural effects of treatments for depression and biomarkers of response: a systematic review of neuroimaging studies. *Psychol Med* 50:187–209. <https://doi.org/10.1017/S0033291719003660>
155. Kang S-G, Cho S-E (2020) Neuroimaging biomarkers for predicting treatment response and recurrence of major depressive disorder. *Int J Mol Sci* 21:2148. <https://doi.org/10.3390/ijms21062148>
156. Baskaran A, Milev R, McIntyre RS (2012) The neurobiology of the EEG biomarker as a predictor of treatment response in depression. *Neuropharmacology* 63:507–513. <https://doi.org/10.1016/j.neuropharm.2012.04.021>
157. Widge AS, Bilge MT, Montana R et al (2019) Electroencephalographic biomarkers for treatment response prediction in major depressive illness: a meta-analysis. *Am J Psychiatry* 176:44–56. <https://doi.org/10.1176/appi.ajp.2018.17121358>
158. Kircanski K, Williams LM, Gotlib IH (2019) Heart rate variability as a biomarker of anxious depression response to antidepressant medication. *Depress Anxiety* 36:63–71. <https://doi.org/10.1002/da.22843>