Journal of Affective Disorders 405 (2026) 121674

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

Development of a baseline battery to optimize depression treatment assignment in primary care: Balancing breadth and brevity

Julia R. Pozuelo [a,b,c,*], Ronald C. Kessler [d], Vikram Patel [a,e], Anuja Lahiri [f], Yashika Parashar [a], Mohammad M. Herzallah [g,h], Mimansa Khanduri [f], Kathryn O'Neill [a], Alyssa Martinez [i], Anant Bhan [f], Daisy R. Singla [j,k,l], John A. Naslund [a], Pim Cuijpers [m,n], Jordan W. Smoller [o,p,q], Karmel W. Choi [o,p], Tyler J. VanderWeele [e], Rahul S.P. Singh [f], Akanksha Shukla [f], Arvind Kushwah [f], Azaz Khan [f], Sruthi G. [f], Varun Shende [f], Yashwant K. Mehra [f], Michelle Melwyn Joel [r], Shubham Atal [r], Abhijit Rozatkar [r], Tamonud Modak [r], Steven D. Hollon [s]

[a] Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA
[b] Department of Psychiatry, University of Oxford, Oxford, United Kingdom
[c] MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[d] Department of Health Care Policy, Harvard Medical School, Boston, MA, USA
[e] Harvard T.H. Chan School of Public Health, Boston, MA, USA
[f] Bhopal Hub, Sangath, Bhopal, India
[g] Palestinian Neuroscience Initiative, Al-Quds University, Abu Dis, Jerusalem, Palestine
[h] Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New Jersey, Newark, NJ, USA
[i] School of Social Work, University of North Carolina at Chapel Hill, North Carolina, USA
[j] Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Canada
[k] Department of Psychiatry, Temerty Faculty of Medicine, University of Toronto, Toronto, Canada
[l] Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada
[m] Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, the Netherlands
[n] International Institute for Psychotherapy, Babeș-Bolyai University, Cluj-Napoca, Romania
[o] Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[p] Psychiatric & Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[q] Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[r] All India Institute of Medical Sciences (AIIMS) Bhopal, Bhopal, India
[s] Department of Psychology, Vanderbilt University, Nashville, TN, USA

ARTICLE INFO

Keywords:
Precision mental health
Depression
Instrument development
Cross-cultural research
Primary care

ABSTRACT

Background: Psychotherapy and antidepressant medications are effective treatments for depression on average, yet individual response varies widely. Predicting which treatment is most effective for a given patient requires a comprehensive yet feasible assessment of baseline characteristics. This paper describes the development of a baseline battery to guide treatment allocation between behavioral activation and fluoxetine for adults with depression in primary care in India.

Methods: We used a six-step, multimethod approach to develop the battery. We updated a prior review to identify constructs associated with differential treatment response, conducted an expert survey (220 invited; 80 responses) to refine the list, and selected measurement tools based on brevity, validity, and scalability. We applied semantic analysis to identify redundancies, piloted the battery with 200 primary care patients, and used advanced statistical methods to optimize item selection.

Results: The final battery comprised 68 constructs spanning clinical, psychological, cognitive, socioeconomic, and biological domains, assessed through questionnaires, tasks, and biomarkers. Pilot testing revealed comprehension challenges, participant fatigue, and low-variability items. Data-driven item reduction removed more than

* Corresponding author at: Department of Global Health and Social Medicine, Harvard University, 641 Huntington Avenue, Boston, Massachusetts, 02115, USA.
E-mail address: julia_ruizpozuelo@hms.harvard.edu (J.R. Pozuelo).

https://doi.org/10.1016/j.jad.2026.121674
Received 3 December 2025; Received in revised form 13 March 2026; Accepted 21 March 2026
Available online 22 March 2026

25% of items while retaining full construct coverage, reducing administration time from over three hours to approximately two.
Conclusion: This study shows how a structured, data-driven development process can yield a scientifically robust baseline assessment battery capturing a broad range of candidate treatment moderators. The battery will undergo further refinement following the results of the OptimizeD trial, which will identify the subset of baseline characteristics that meaningfully predict differential treatment response. While not intended as an implementation-ready clinical tool, the development framework presented here may help guide other researchers conducting similar precision treatment studies, particularly in resource-constrained settings.

1. Introduction

Major depressive disorder is one of the most common psychiatric disorders and a leading cause of disability worldwide (Ferrari et al., 2024; McGrath et al., 2023). In primary care, two commonly recommended first-line treatments are behavioral activation (a brief evidence-based psychotherapy) and antidepressant medication, most commonly selective serotonin reuptake inhibitors (SSRIs) such as fluoxetine (Patel et al., 2016; WHO, 2010). While these interventions show comparable average treatment effects, individual response varies widely (Cuijpers et al., 2013, 2020; Maslej et al., 2021; Kaiser et al., 2022). This variability has prompted efforts to develop precision treatment approaches that match patients to the treatment most likely to be effective for them.

A key distinction in these efforts lies between prognostic and prescriptive indicators. *Prognostic* indicators predict overall response to treatment, identifying patients who are more or less likely to improve regardless of the intervention they receive. In contrast, *prescriptive* indicators identify characteristics that predict whether a patient is more likely to benefit from one type of treatment than another. While both are useful, most existing research has focused on prognostic indicators, with relatively little attention given to prescriptive factors that can directly guide treatment selection (Cohen and DeRubeis, 2018). Early efforts to match patients to treatments relied on single indicators, such as depression subtype and severity, but no individual characteristic has shown sufficient predictive power to inform clinical decisions meaningfully (Cuijpers et al., 2012; DeRubeis et al., 2014; Kessler, 2018; Kessler et al., 2017). More recently, advances in machine learning have enabled the integration of multiple patient characteristics to improve personalized treatment assignment (Lee et al., 2018; Ermers et al., 2020; Chekroud et al., 2021; Puac-Polanco et al., 2023; LoParo et al., 2025; Chekroud et al., 2016). However, the performance and generalizability of these models depend critically on the quality, breadth, and feasibility of the baseline data on which they are trained. While many potential predictors – both prognostic and some prescriptive – have been proposed (Cohen and DeRubeis, 2018; Fournier et al., 2009; Kemp et al., 2008; Kessler et al., 2017), selecting which constructs to measure and how best to assess them remains a key challenge. Long and burdensome assessments reduce data quality and limit clinical scalability, underscoring the need for streamlined tools that retain predictive utility while remaining feasible for use in real-world settings.

In addition to challenges in measurement, several broader limitations continue to hinder progress in precision treatment research. First, small sample sizes in much of the existing literature limit statistical power to detect treatment-by-patient interactions and undermine the reliability of predictive models (Luedtke et al., 2019). Second, studies often assess only a limited number of patient characteristics, such as genetic markers, personality traits, or socioeconomic status, typically in isolation rather than as part of an integrated, cross-disciplinary framework (Bucher et al., 2019; De Carlo et al., 2016; Fonseka et al., 2018; Mulder, 2002; Newton-Howes et al., 2014; Phillips et al., 2015). Third, while there have been head-to-head comparisons of antidepressants and psychotherapy, much of the literature evaluates each against nonspecific controls, offering limited insight into prescriptive factors that can inform differential treatment selection (Cuijpers et al., 2013). Fourth, most precision treatment studies have been conducted in secondary or tertiary care settings, despite primary care being the first point of contact for most patients. Primary care populations differ from those in specialty care in ways that may affect treatment response, including milder symptom severity, greater comorbidity and stigma, and smaller effect sizes (Cuijpers et al., 2023; Driessen et al., 2010; Fournier et al., 2010). Finally, the majority of this work has been conducted in high-income countries, which limits its relevance for low-resource health systems, where scalable and contextually appropriate predictive models are particularly needed (Moitra et al., 2022; Rathod et al., 2017).

The Optimizing Depression (OptimizeD) trial aimed to address these challenges by developing a comprehensive baseline battery for use in primary care that can predict whether an individual patient is more likely to benefit from fluoxetine or behavioral activation (Pozuelo et al., 2025). Both treatments have demonstrated effectiveness across a range of international settings, including specifically in India, where the current study was conducted (Patel et al., 2017, 2003). This paper describes the multistep process undertaken to develop the OptimizeD baseline battery, focusing on strategies to streamline the assessment while maintaining the breadth across constructs needed to identify meaningful prescriptive indicators. The battery is intended as a research-stage discovery instrument and methodological demonstration: it is deliberately comprehensive to maximize the chance of identifying prescriptive predictors, with the expectation that trial results will enable substantial shortening toward a scalable clinical tool. This is made possible by the scale of the OptimizeD trial, which randomizes 1500 participants – well above the 600–1000 total estimated to be required to develop a useful algorithm (Pozuelo et al., 2025). While focused on a comparison of two specific treatments in one setting, the resulting battery provides a practical and generalisable framework for advancing precision treatment research in low-resource primary care settings, and may serve as a template for future precision studies in depression treatment.

2. Methods

We used a six-step, systematic approach to develop the baseline battery, including updating a previous review, surveying experts, selecting brief and scalable measures, assessing redundancy, pilot testing in Indian primary care settings, and applying machine learning to streamline the tool. These steps are summarized in the sections that follow. All study procedures involving human subjects received ethics approval from institutional review boards in India (Sangath IRB: AB-2021-69; AIIMS Bhopal: EF0237) and the United States (Harvard Medical School IRB20–2144).

2.1. Step 1: Update of previous review

To identify potential predictors of differential treatment response, we first updated a previous systematic review by Kessler and colleagues that examined self-reported measures associated with differential treatment response (Kessler et al., 2017). The original review identified 54 studies that met the inclusion criteria and reported on 28 baseline predictors of treatment response across six domains (sociodemographic, history of illness, comorbidities, life stress, personality traits, and others). The methodology for our update closely followed that of the original review. We used the same database (PubMed), search strategy

(["depress*" AND "predict*" AND ("treatment outcome" OR "treatment response" OR "course") AND ("self-report" OR "survey" OR "questionnaire")]*), and inclusion criteria. Eligible studies involved participants receiving treatment for major depressive disorder (whether in randomized controlled trials, uncontrolled trials, or observational studies) and examined associations between baseline constructs and treatment outcomes. Our inclusion criteria did not restrict studies to fluoxetine (the SSRI used in OptimizeD), as our goal was to identify candidate predictors of antidepressant response broadly. We therefore included studies examining a range of pharmacological treatments – including agents less commonly used in primary care – given evidence that certain predictors operate across multiple antidepressant agents rather than being entirely drug-specific (Chekroud et al., 2017).

The literature search was conducted on September 9, 2022, and was restricted to English-language articles published between 2014 and 2022. A single researcher (AM) compiled the search results and removed duplicates. Two reviewers (AM, SDH) independently screened the titles and abstracts of all retrieved references using Covidence to identify potentially eligible studies (Veritas Health Innovation, 2024). Full texts were obtained for studies that met the inclusion criteria or required further evaluation. Two reviewers (AM, KO) then independently assessed full texts for final inclusion and extracted data from the included studies, resolving any disagreements through discussion and consensus. To ensure consistency and accuracy, a third reviewer (JRP) subsequently reviewed all full-text articles. Reasons for excluding studies after full-text review were documented. Study quality was not assessed as part of this review update.

2.2. Step 2: Obtaining expert opinion

We conducted an expert survey to refine our list of potential predictors and identify overlooked constructs. The survey targeted lead and senior authors from studies included in the reviews, as well as additional experts identified through hand-search and peer recommendations. Experts were asked to: (i) rate each construct on its likelihood of predicting treatment response to behavioral activation and antidepressant medication; (ii) suggest additional constructs that were not included in the survey but might be relevant predictors; and (iii) recommend other experts who should be consulted in this process. Experts received a Qualtrics survey link via email on December 16, 2022 (the survey can be accessed via this link: https://hms.az1.qualtrics.com/jfe/form/SV_baxdzDpQui8Gyy2). The survey was sent to 220 experts. Non-respondents received a single reminder email. The survey contained 51 constructs derived from the reviews, and experts rated them on a binary scale ("Unlikely" = 0 vs. "Likely" = 1) in terms of their predictive value for each treatment modality separately. For each construct and treatment modality, we calculated the mean rating across experts. Given the binary response format, the mean corresponds to the proportion of experts endorsing the construct as "likely" to predict treatment response to behavioral activation or antidepressant medication and therefore serves as an index of expert consensus. Values closer to 1 indicate broad agreement that a construct is prognostic, values closer to 0 indicate agreement that it is not, and intermediate values reflect greater divergence in opinion. Open-ended fields allowed respondents to suggest additional constructs and recommend other experts to consult. Additionally, all newly suggested constructs were added to our list of predictors for further evaluation.

2.3. Step 3: Selection of measurement tools

After finalizing the list of predictors based on the systematic review (Step 1) and expert survey (Step 2), we identified appropriate measurement tools for each construct to be assessed via questionnaires. For each construct (e.g., sleep problems), we reviewed multiple candidate instruments, including brief items embedded within broader depression scales (e.g., PHQ-9, Hamilton Depression Rating Scale) as well as stand-alone measures (e.g., Insomnia Severity Index, Pittsburgh Sleep Quality Index). Available tools were systematically identified and compared against four pre-specified criteria: (i) *Brevity:* Preference was given to measures with fewer than 10 items to minimize participant burden and enhance usability in primary care settings; (ii) *Prior validation:* Tools with existing validation in India or similar low-resource settings were prioritized to improve cultural and contextual appropriateness; (iii) *Language availability:* Measures available in Hindi were favored to facilitate implementation without the need for additional translation and validation; and (iv) *Public accessibility:* Open-access and non-proprietary tools were prioritized to reduce costs and enhance scalability for future use. We identified and compared available tools for each construct to determine which met the highest number of desirable criteria. We applied a similar criterion to identify neurocognitive tasks for each construct, prioritizing tools that were brief, freely available, and previously validated in comparable settings.

2.4. Step 4: Assessing redundancy across tools

Many of the selected measurement tools included conceptually similar or nearly identical items, resulting in redundancy and increasing the length of the battery without improving construct coverage. For example, several scales included items assessing sleep problems or headaches, often worded slightly differently but evaluating the same underlying construct. To address this, we calculated an index of item semantic similarity to detect overlapping content across different measurement scales. We computed pairwise semantic similarity scores across the selected measurement tools for all items. This analysis was conducted in both English and Hindi to ensure cross-linguistic consistency. All items were extracted from the selected measures and pre-processed to ensure standardization. This involved removing extraneous text, such as response scale instructions and numerical labels, to isolate item content for comparison. Next, every item was encoded as a high-dimensional vector representation using OpenAI's text-embedding-ada-002 embedding model (OpenAI, 2023). This model captures the position of the tokenized text in the language vector space. Pairwise cosine similarity scores (normalized vector dot products) were then calculated between all items to assess their degree of semantic relatedness. Pairs with a similarity score of 90% or higher were flagged as potential duplicates. Given the large number of total item pairs (>67,00), we focused our manual review on the 1% with the highest similarity scores. Two independent raters (AL, YP) reviewed these pairs in both English and Hindi to validate the automated findings. Each pair was assigned a similarity score using a standardized rating system (0 = totally separate, 0.5 = some overlap, and 1 = perfect agreement). Inter-rater reliability was assessed using Cohen's kappa to measure agreement between raters.

2.5. Step 5: Pilot testing

To assess the feasibility and appropriateness of the selected measures, we conducted a pilot study in Bhopal, Madhya Pradesh, a state in central India with a population of over 87 million, of whom nearly 73% reside in rural areas (Gov. of India, 2011; World Population, 2024). The pilot study aimed to evaluate the clarity, cultural relevance, and variability of the measures to ensure that the final battery was well-suited for use in primary healthcare settings. This pilot included the full baseline battery (comprising questionnaires and neurocognitive tasks) while blood sample collection was conducted with a subset of 76 participants to assess the feasibility and acceptability of biological data collection.

Participants were recruited from eight Primary Healthcare Facilities in Bhopal and included adults aged 18 years or older, both with and without depression, as measured by the Patient Health Questionnaire (PHQ-9) using a cutoff of 10 (Kroenke et al., 2001). The PHQ-9 and the cutoff score of 10 have been validated in India, demonstrating good sensitivity and specificity for detecting major depression in primary care

settings (Patel et al., 2008). Participants were recruited regardless of whether they were seeking treatment for depression, ensuring a sample that reflects the broader primary care population rather than the treatment-seeking population typically seen in secondary care. This design ensured a full range of scores, allowing us to assess the sensitivity of the measures across different symptom levels. In addition to ethics approval, permission was obtained from local health authorities to recruit participants from primary care clinics. All participants provided written informed consent or verbal (witnessed) consent for participants who were illiterate, per local regulations (and as approved by the ethical review boards at each institution). Participants were provided with INR 500 (~US$6) as compensation for their time for completing the baseline survey.

We collected both qualitative and quantitative data to identify problematic questions and refine measures. Qualitative feedback was gathered using two approaches. First, interviewers completed a post-fieldwork survey at the end of each day, documenting any difficulties encountered during data collection. Second, we audio-recorded assessments (with participant consent) and a research assistant (YP) reviewed and coded these interviews to identify participant difficulties, interviewer challenges, and contextual barriers. Together, these data sources provided a comprehensive understanding of field implementation challenges and informed the refinement of measures. Questions were flagged if they were frequently misread by interviewers, required repeated clarification from participants, or made participants uncomfortable. For the quantitative analysis, we inspected the data to identify potential measurement issues. For example, we flagged overly lengthy scales that contributed to participant fatigue and items with low variability, to which most participants selected the same response, indicating limited discriminative power.

2.6. Step 6: Item reduction analyses

Lastly, to reduce participant burden while maintaining coverage of each construct, we conducted a two-stage item reduction process for each construct using data from participants with moderate to severe depression (PHQ-9 $\geq$ 10). Analyses were conducted in this subsample to ensure that item selection reflected variation within the main trial's target population, rather than being influenced by responses from non-depressed participants for whom many constructs may have limited applicability. Single-item constructs (e.g., history of illness, treatment preference) were excluded from this process due to their brevity and relevance, as were sociodemographic characteristics, which were collected to characterize the sample. The analysis was conducted using Stata (version 17) (StataCorp, 2021).

In the first stage, we performed a principal-component factor analysis to identify items that contributed weakly to the scale. We first examined the scree plots to determine the appropriate number of factors and tested one- and two-factor solutions. When a two-factor solution was indicated, Promax rotation was applied to allow for correlated factors (Fabrigar et al., 1999). Items with absolute factor loadings <0.4 were considered weak contributors and examined for removal. Constructs that did not align with theoretically expected factor structures (e.g., unidimensional constructs splitting into multiple factors) were reviewed. After finalizing item selection, a total score was computed for each construct using the retained items. In the second stage, we applied lasso regression with 10-fold cross-validation to refine the selection of the most predictive items of total scale scores (Tibshirani, 1996). Lasso penalizes less important predictors, reducing redundancy while ensuring optimal predictive value in the context of the linear additive link function. We examined coefficient paths to evaluate how variables entered or were excluded at different penalty levels.

3. Results

3.1. Step 1: Update of previous review

Our literature search identified 2797 potentially relevant studies. Removing one duplicate left 2796 unique records for title and abstract screening. Of these, 2679 studies were excluded based on relevance, leaving 117 full-text articles for further review. Following full-text screening, an additional 51 studies were excluded, resulting in a final inclusion of 66 studies (see Fig. 1).

Most of the included studies were conducted in high-income countries, with the United States (42%), Canada (12%), and Germany (11%) being the most frequently represented. They were primarily randomized controlled trials (50%) and cohort studies (38%) evaluating the effectiveness of antidepressant medications (45%), psychological therapies (44%), or combination treatments (6%). Table 1 summarizes the key characteristics of these studies, while Table A1 in the supplementary materials details excluded studies and their reasons for exclusion.

Through this updated review, we identified 23 additional constructs, expanding the total number of constructs from 28 in Kessler et al. (2017) to 51 in our combined review. This final set of 51 constructs was subsequently presented to experts for evaluation.

3.2. Step 2: Obtaining expert opinion

Of the 220 total experts invited, 80 (36%) completed the survey. The average time to complete the survey was 17.9 min (95% CI: 12.4–23.5 min). Most respondents were based in institutions in high-income countries, with the largest representation from the United States (36%) and the United Kingdom (25%). Only two respondents were based in middle-income countries (one from China and one from India). This geographical distribution is consistent with the fact that our expert sample was derived from authors of papers included in both reviews, which predominantly comprised studies conducted in high-income settings. Respondents represented a broad range of institutions and disciplines (Supplementary Table A2). The majority were senior academics (primarily Professors and Associate Professors) with extensive research experience, spanning clinical psychology, psychiatry, psychiatric epidemiology, and related fields (mean citations = 42,249; SD = 79,906). Although institutional affiliations were concentrated in high-

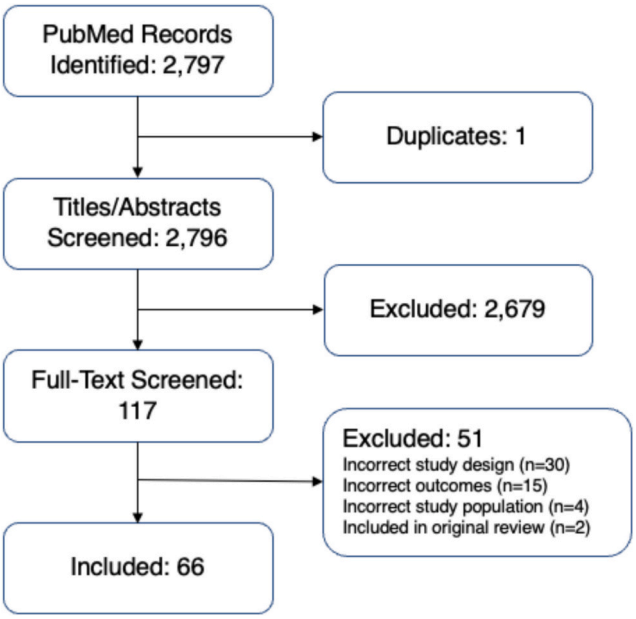

Fig. 1. Study selection.

Table 1
Characteristics of included studies.

#	Citation	Country	Study Type	Intervention	**Comparator**	**Constructs**
1	(Akechi et al., 2019)	Japan	RCT	Pharmacotherapy (mirtazapine)	Pharmacotherapy (sertraline or sertraline augmented with mirtazapine)	Depression severity, suicidal ideation, medication side effects
2	(Allen et al., 2021)	Canada	RCT	Pharmacotherapy (escitalopram augmented with aripiprazole)	Pharmacotherapy (escitalopram)	Neuroticism, stressful life events
3	(Arnow et al., 2015)	United States, the Netherlands, Australia, New Zealand, and South Africa	RCT	Pharmacotherapy (Escitalopram or Sertraline)	Pharmacotherapy (Venlafaxine)	Depression subtype (melancholic, atypical, anxious) [not significant]
4	(Bauer et al., 2015)	Australia, Belgium, Canada, Czech Republic, Finland, France, Germany, Norway, Poland, Romania, South Africa, Sweden, USA.	Pooled analysis of two RCTs	Pharmacotherapy (extended-release quetiapine fumarate)	Placebo	Baseline psychiatric history, demographic and disease characteristics (authors found no predictive association between constructs and efficacy outcomes)
5	(Bausch et al., 2017)	Germany	RCT	Psychological (Cognitive Behavioral Analysis System of Psychotherapy)	Pharmacotherapy (escitalopram plus clinical management)	Childhood maltreatment, depression severity
6	(Békés et al., 2015)	USA	Observational study	Depression treatment (any; 87% were on medication)	N/A	Perfectionism (self-criticism, personal standards, neuroticism, conscientiousness) and interpersonal chronic stres
7	(Buckman et al., 2019)	United Kingdom	Observational study	Psychological intervention (low & high intensity therapies)	N/A	Attention control
8	(Chekroud et al., 2017)	USA	Pooled analysis of 9 RCTs	Pharmacotherapy (citalopram hydrobromide, escitalopram plus monotherapy, escitalopram plus bupropion hydrochloride, venlafaxine hydrochloride plus mirtazapine)	Placebo or Pharmacotherapy (duloxetine)	Three derived symptom clusters (emotional, sleep, atypical/somatic) found to be predictive of differential treatment response
9	(Chui et al., 2016)	USA	RCT	Pharmacotherapy (sertraline, venlafaxine) & psychological intervention (supportive-expressive therapy)	Placebo	Self-criticism, dependency, connectedness
10	(Cohen et al., 2020)	Netherlands	RCT	Psychological intervention (CBT)	Psychological intervention (psychodynamic therapy)	Depression severity, anxiety sensitivity, extraversion, and psychological treatment-needs
11	(Constantinou et al., 2020)	USA	Observational study	Psychological (6–8 week course of multimodal inpatient treatment)	N/A	Personality disorder symptoms
12	(Dawson et al., 2017)	USA	Observational study	Psychotropic medication only, psychotherapy only, or a combination of medication and therapy	N/A	Executive Functioning
13	(Dinger et al., 2017)	Germany	RCT	Psychological intervention (multimodal, psychodynamically oriented, day-clinic)	Psychological intervention (multimodal, psychodynamically oriented, inpatient)	Self-esteem
14	(Dunlop et al., 2020)	Canada	Non-randomized, open-label clinical trial	Pharmacotherapy (escitalopram)	Healthy controls	Reward processing
15	(Edmonds et al., 2018)	Canada	Observational study	Psychological intervention (internet-delivered CBT)	N/A	Baseline severity, education level, being on disability
16	(Fitzpatrick et al., 2020)	Not specified	Observational study	Psychological intervention (CT)	N/A	Maladaptive personality traits, interpersonal problems, social skills
17	(Forand and DeRubeis, 2014)	Not specified	RCT	Responders to psychological intervention (CT)	Responders to pharmacotherapy with paroxetine (could be augmented with lithium or desipramine).	Extreme response style / Dysfunctional Attitude
18	(Fowler et al., 2015)	USA	Observational study	Pharmacotherapy, Psychological Intervention	N/A	Early treatment (non) response
19	(Furukawa et al., 2020)	Japan	RCT	Pharmacotherapy (sertraline and mirtazapine)	Pharmacotherapy (switching to mirtazapine)	*Socio-demographic variables* (age, sex, education, employment status, and marital status), *clinical characteristics* (age at onset, number of previous depressive episodes, length of episode, concurrent physical illness), and *depression chara*cteristics

(*continued on next page*)

Table 1 (*continued*)

#	Citation	Country	Study Type	Intervention	**Comparator**	**Constructs**
						during treatment (severity, adherence to pharmacotherapy).
20	(Gaspersz et al., 2017)	Netherlands	Observational study	Pharmacotherapy	N/A	Anxious distress
21	(Gerra et al., 2014)	USA	RCT	Pharmacotherapy (bupropion extended-release or escitalopram)	Pharmacotherapy (bupropion extended-release and escitalopram)	Negative affectivity
22	(Gollan et al., 2016)	USA	Observational study	Psychological intervention (behavioral activation)	Healthy controls	Negativity bias (significant), positivity offset (not significant)
23	(Hallgren et al., 2017)	Sweden	RCT	Psychological intervention (exercise or CBT)	Treatment as usual	Social relationships
24	(Hallgren et al., 2016)	Sweden	RCT	Psychological intervention (CBT)	Treatment as usual	Habitual physical activity levels
25	(Hornstein et al., 2021)	Not specified	Observational study	Psychological intervention (CBT, mindfulness, psychoeducation, heart rate variability biofeedback)	N/A	Baseline depression severity, baseline anxiety severity, self-reported motivation, type of referral into the programme (self vs. healthcare provider), Work Productivity and Activity Impairment
26	(Jha et al., 2018)	USA	RCT	Pharmacotherapy (bupropion-plus-escitalopram, venlafaxine-plus-mirtazapine)	Pharmacotherapy (escitalopram monotherapy)	Sub-threshold hypomanic symptoms
27	(Jha et al., 2016)	USA	RCT	Pharmacotherapy (bupropion-plus-escitalopram, venlafaxine-plus-mirtazapine)	Pharmacotherapy (escitalopram plus placebo)	Improvement in work productivity
28	(Kaneriya et al., 2016)	Not specified	RCT	Pharmacotherapy (venlafaxine and aripiprazole)	Pharmacotherapy (venlafaxine and placebo)	Executive function, severity of anxiety, and severity of medical comorbidity (not significant)
29	(Keefe et al., 2016)	Not specified	Secondary analysis of RCT (intervention arm only)	Psychological intervention (CT)	N/A	Increased focus on maladaptive beliefs, personality disorders
30	(Kertz et al., 2015)	USA	Observational study	Psychological intervention (CBT) and pharmacotherapy	N/A	Repetitive negative thinking
31	(Khazanov et al., 2020)	USA	RCT	Combined pharmacotherapy and psychological intervention (CT)	Pharmacotherapy (antidepressant-only)	Distress and anhedonia
32	(Kudo et al., 2016)	Japan	Observational study	Naturalistic, uncontrolled treatment	N/A	Personality traits (significant: personal reserve, rejection sensitivity, and self-criticism; not significant: rejection sensitivity and self-criticism)
33	(Laird et al., 2018)	USA	RCT	Pharmacotherapy (methylphenidate or citalopram)	Pharmacotherapy (methylphenidate and citalopram)	Resilience (grit, active coping self-efficacy, accommodative coping self-efficacy, and spirituality)
34	(Lorenzo-Luaces et al., 2020)	USA	RCT	Psychological intervention (CBT), Pharmacotherapy (fluoxetine) alone and in combination	Placebo	Depression severity and chronicity
35	(Menchetti et al., 2014)	Italy	RCT	Psychological (interpersonal counseling)	Pharmacotherapy (sertraline or citalopram)	First episode of depression, baseline depression scores, daily functioning, comorbidity with anxiety disorder, smoking
36	(Metts et al., 2018)	USA	Observational study	Psychological intervention (CBT)	N/A	Neuropsychological functioning
37	(Mihaljevic et al., 2016)	Croatia	Observational study	Pharmacotherapy (SSRI and benzodiazepines)	N/A	Religiousness and spirituality
38	(Mills et al., 2022)	USA	RCT	Pharmacotherapy (escitalopram plus bupropion or venlafaxine plus mirtazapine)	Pharmacotherapy (escitalopram plus a placebo)	Education level, employment status, minority status
39	(Moradveisi et al., 2015)	Iran	RCT	Psychological (behavioral activation)	Pharmacotherapy (sertraline)	Attribution of treatment effects (patients' beliefs about why they recovered)
40	(Nguyen et al., 2022)	USA	RCT	Pharmacotherapy (sertraline; sertraline nonresponders switched to bupropion)	Placebo	Reward processing
41	(Perlis et al., 2014)	USA	Prospective, multi-step treatment trial (analysis of the STAR*D)	Pharmacotherapy (citalopram and augmentation with buspirone, bupropion or CT, or switch to venlafaxine, sertraline, bupropion or cognitive therapy)	Pharmacotherapy (citalopram)	Manic/hypomanic symptoms (DSM-5 mixed features)
42	(Perlman et al., 2019)	Not specified	Systematic meta-review	Pharmacotherapy (any)	N/A	Sociodemographic, symptom profile, genetic, neuroimaging,

(*continued on next page*)

Table 1 (*continued*)

#	Citation	Country	Study Type	Intervention	**Comparator**	**Constructs**
						clinical comorbidities, and peripheral markers. But no conclusions about the overall directionality of each.
43	(Quilty et al., 2017)	Canada	Observational study	Pharmacotherapy (high versus a low affinity for the serotonin transporter)	N/A	Childhood abuse
44	(Radkovsky et al., 2014)	Germany	Single arm trial	Psychological (individual & group CBT)	N/A	Emotion regulation skills
45	(Rnic et al., 2021)	Canada	Non-randomized, open-label clinical trial	Pharmacotherapy (escitalopram)	N/A	Cognitive self-appraisals
46	(Rossom et al., 2016)	USA	Observational study	Usual care for depression	N/A	Health status, employment status, age, depression severity
47	(Saunders et al., 2016)	United Kingdom	Observational study	Stepped-care approach, with wide range of psychological treatments	N/A	Age at referral, gender, depressive symptoms, anxiety symptoms, functioning, merdication prescription status, welfare status, ethnic group, phobia
48	(Schilling et al., 2014)	Germany	Observational study	Psychological (inpatient multimodal psychotherapy)	N/A	Childhood abuse and neglect
49	(Schilling et al., 2021)	Germany	Secondary analysis of RCT (intervention arm only)	Psychological (CBT)	N/A	Life events, motivation, and risk/suicidality
50	(Stikkelbroek et al., 2020)	Netherlands	RCT	Psychological (CBT)	Treatment as usual	Age, gender, child/parent educational level, suicidal criteria, comorbidity, and severity of depression (none of them were significant)
51	(Stiles-Shields et al., 2015)	USA	RCT	Psychological (CBT over telephone)	Psychological (CBT face-to-face)	Self-efficacy, baseline depression severity, education, social support, physical functioning, and employment status, sex
52	(Stiles-Shields et al., 2014)	USA	RCT	Psychological (CBT over telephone)	Psychological (CBT face-to-face)	Comorbid anxiety
53	(Sumiyoshi et al., 2021)	Japan	Observational study	Pharmacotherapy (a SSRI, SNRI, NaSSA and/or TCA/TeCA)	N/A	Cognitive impairment, psychosocial function, quality of life
54	(Tait et al., 2022)	United Kingdom	Observational study	Psychological (CBT, counseling for depression, interpersonal psychotherapy, dynamic interpersonal therapy)	N/A	Coping strategies (engagement coping and rumination) [not significant]
55	(Thiruchselvam et al., 2019)	Canada	RCT	Psychological (CBT)	Pharmacotherapy (Bupropion, sertraline, venlafaxine, citalopram, escitalopram, fluoxetine, mirtazapine, and/or duloxetine)	Outcome/change expectancy in depression
56	(Vinckier et al., 2017)	France	Observational study	Pharmacotherapy (agomelatine)	N/A	Anhedonia
57	(Wagner et al., 2017)	Germany	RCT	Pharmacotherapy (escitalopram + early medication change to venlafaxine followed by an augmentation with lithium after non-response)	Pharmacotherapy (escitalopram for additional 2-weeks, followed by venlafaxine in case of non-response)	Depression severity, suicidality, depression subtype, comorbidities (social phobia, avoidant personality disorder)
58	(Wallert et al., 2022)	Sweden	Observational study	Psychological (internet-based CBT)	N/A	Demographic (age, education, work experience), clinical (prior episodes/treatment, severity, quality of life, substance, social anxiety, panic, alcohol use, sleep, sex drive, facial expressions, agitation, functioning), process (e.g., time to complete online questionnaires), and genetic (polygenic risk scores)
59	(Webb et al., 2020)	USA	Observational study	Psychological (CBT group and individual) and pharmacotherapy	N/A	Depression severity, fatigue, concentration, anxiety severity, relationship problems, personality disorder, expectations of improvement, prior treatment, age of onxet, social anxiety, post-traumatic

(*continued on next page*)

Table 1 (*continued*)

#	Citation	Country	Study Type	Intervention	**Comparator**	**Constructs**
						stress disorder, mood stabilizer prescription, race
60	(Whiston et al., 2019)	Not specified	Systematic review & data analysis	Psychological (CBT)	Psychological (Interpersonal psychotherapy)	Age, depression severity, individual format of administration, no adjuntive ADMs
61	(Williams et al., 2016)	United States, Netherlands, Australia, New Zealand, South Africa	RCT	Pharmacotherapy (escitalopram, sertraline or venlafaxine)	Pharmacotherapy (escitalopram, sertraline or venlafaxine)	Childhood trauma
62	(Worley et al., 2014)	USA	Secondary analysis of RCT	Psychological (Integrated CBT)	Psychological (Twelve-Step Facilitation)	Neurocognitive impairment
63	(Yun et al., 2020)	Canada	Observational study	Pharmacotherapy (vortioxetine)	N/A	Early increases in physical activity
64	(Zilcha-Mano et al., 2019)	USA	Data combined from two RCTs	Pharmacotherapy (citalopram or escitalopram	Placebo	Patient expectancy
65	(Ziobrowski et al., 2022)	USA	Observational study	Psychological (therapy not specified)	N/A	Antecedent and concomitant psychiatric conditions, clinical staging, treatment characteristics, protective factors/resilience, and socio-demographics
66	(Zisook et al., 2019)	USA	RCT	Pharmacotherapy (switch to another antidepressant: sustained-release bupropion; combination with another antidepressant: sustained-release bupropion; or augmentation with an antipsychotic—aripiprazole)	Pharmacotherapy	Employment status, severity, chronicity, anxiety, grief, childhood adversity, quality of life/positive mental health, age, mixed hypomanic symptoms

Note. Antidepressant medication (ADM), randomized controlled trial (RCT), cognitive behavioral therapy (CBT), cognitive therapy (CT), Twelve-Step Facilitation, selective serotonin reuptake inhibitor (SSRI), serotonin–norepinephrine reuptake inhibitor (SNRI), noradrenergic and/or specific serotonergic antidepressant (NaSSA), tricyclic/tetracyclic antidepressant (TCA/TeCA).

income settings, more than half had prior research experience in low- and middle-income countries.

Experts provided ratings for constructs predicting treatment outcomes separately for either antidepressant medication or behavioral activation. The constructs evaluated spanned multiple domains, including sociodemographic factors, history of illness, and personality traits. A total of 28 constructs were rated as having greater prognostic value for behavioral activation, versus 23 constructs for antidepressant medication (Fig. 2). Constructs such as depression severity and chronicity received high endorsement proportions across both treatments, reflecting broad agreement regarding their prognostic relevance. In contrast, constructs such as religion received low endorsement proportions, indicating consensus that they are unlikely to predict treatment response. Other constructs (e.g., age) showed intermediate endorsement, reflecting a lack of strong agreement among experts. Thirty-one of the experts suggested additional constructs, many of which overlapped with each other or with constructs already included in the initial list. Based on these suggestions, we added nine new constructs to the final list, such as grief, markers of inflammation, and patients' treatment outcome expectations (i.e., whether they believe drug treatment or talking therapy would be most effective for them). This process resulted in a final list of 68 constructs: 60 assessed via self-report questionnaires, 5 via neurocognitive tasks, and 3 via biological or genetic indicators. Appropriate tools for each are described in the following section.

3.3. Step 3: Operations (measurement tools) selection

Across the 60 constructs identified in Steps 1 and 2 for questionnaire-based assessment, we reviewed 144 candidate instruments and selected one primary measure per construct (Table 2). The majority of selected tools fulfilled three or more of the four predefined criteria, and a third met all four. The selected instruments were generally brief (median number of items = 5; range = 1–20), supporting feasibility in primary care settings. Most measures were publicly accessible and non-proprietary. However, few were available in Hindi, requiring translation and back-translation procedures for most instruments.

Further, we reviewed several options for assessing cognitive functioning, including the NIH Toolbox Cognition Battery and the ICMR-Neurocognitive Toolbox (Iyer et al., 2020; Weintraub et al., 2013). Ultimately, we selected a subset of neurocognitive tasks from the NIH-funded Longitudinal Aging Study in India – Diagnostic Assessment of Dementia (LASI-DAD), a nationally representative, longitudinal study with extensive cognitive data (Lee et al., 2019). These tasks were previously validated in India, including in Madhya Pradesh, and underwent rigorous translation and cultural adaptation.

3.4. Step 4: Assessing redundancy across tools

To reduce potential redundancy across the 60 selected questionnaire-based constructs, we conducted a semantic similarity analysis on the full set of items. A total of 67,161 possible item-pair combinations were analyzed for semantic similarity in both English and Hindi. Among these, 3734 pairs ranked in the top 5% of most similar items based on their computed similarity scores. Fig. A1 in the supplementary materials shows the semantic similarity matrix for English and Hindi, with yellow indicating the highest similarity and blue representing the most dissimilar items. The correlation between English and Hindi similarity scores was 0.73 (significant at 1%), indicating a strong but imperfect alignment between the two languages in item similarity. Further, there were more items rated as similar in Hindi than in English, as shown by the distribution of similarity scores (Fig. A2 in the supplementary materials). Two human raters independently reviewed the top 1% of pairs (corresponding to 1080 pairs) in both English and Hindi. Despite the high similarity scores assigned by the embedding model, human inspection revealed that many of these item pairs were, in fact, meaningfully different. The model's failures fell into two categories. In some cases, high similarity scores were driven by shared surface structure: items with near-identical wording but different clinical meaning received very high scores despite being conceptually distinct. For

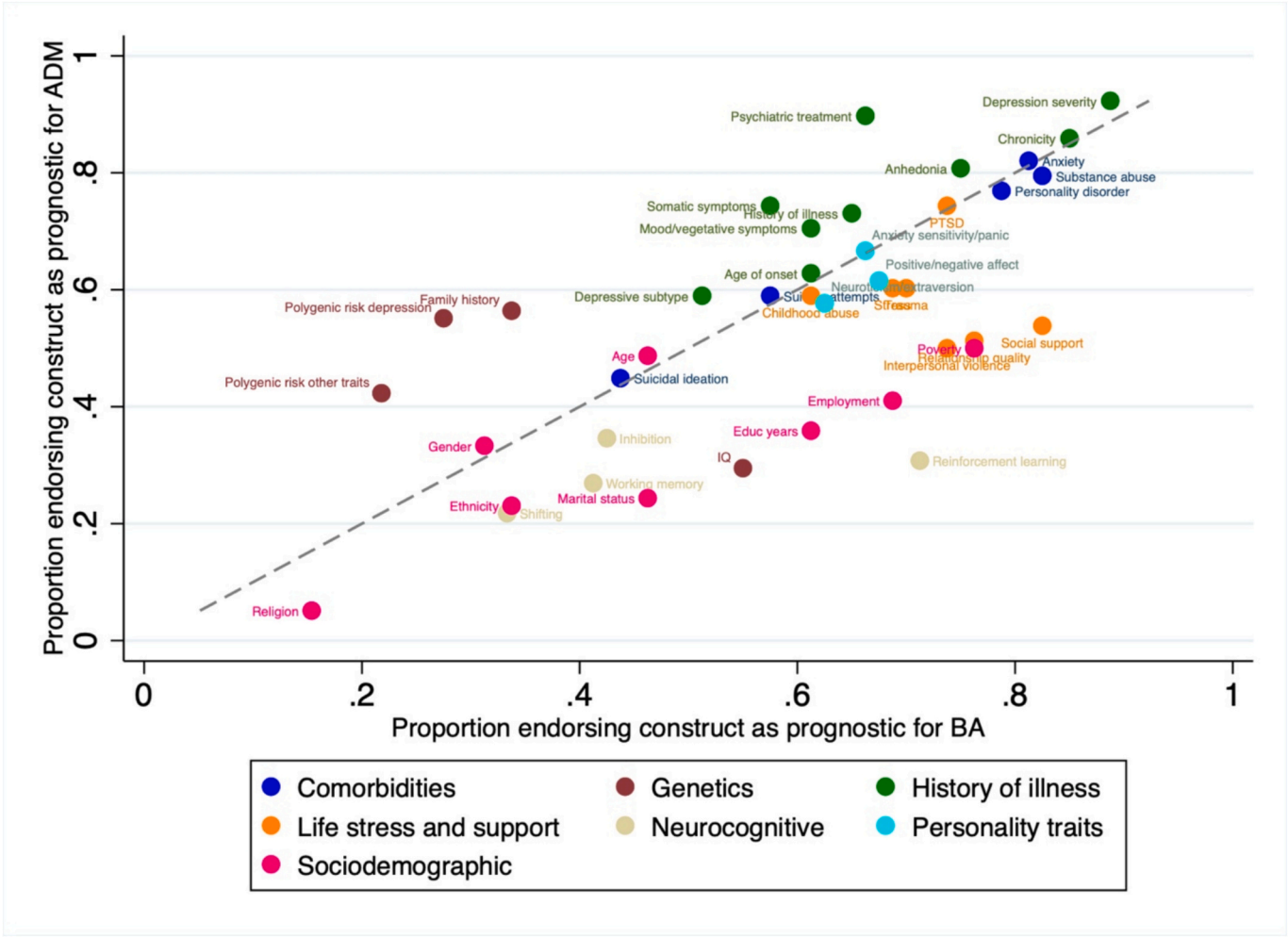

Fig. 2. Proportion of experts endorsing constructs as prognostic for behavioral activation and antidepressant medication
NB. Ratings were binary ("likely" vs. "unlikely"). Points show the proportion endorsing each construct as likely to predict response. Values near 1 indicate strong consensus, values near 0 indicate consensus that it is not prognostic, and intermediate values reflect disagreement. The dashed line denotes equal endorsement across treatments.

example, a pair asking about "past year" versus "lifetime" episodes of low mood scored 0.983, and a fatigue item from the PHQ-9 was paired with a nausea item from the PHQ-15 (0.967); likely because both begin with the introductory phrasing "Over the last two weeks." In other cases, high scores appeared with no obvious structural explanation. For example, an item asking how often a respondent felt overwhelmed in the past month was flagged as highly similar to an item asking about treatment preferences; similarly, an item about bouncing back from hardship was paired with one about keeping emotions to oneself. These failures suggest the model failed to capture meaning and that similarity scores were unreliable indicators of true item redundancy.

Interrater reliability, as assessed by Cohen's kappa, was higher for English ($\kappa = 0.40$) than for Hindi ($\kappa = 0.26$), suggesting that semantic distinctions were easier to recognize in English. The lower agreement in Hindi may reflect translation challenges or cultural variations in item interpretation. Ultimately, this exercise did not help in reducing the number of survey items, as human raters determined that most flagged pairs were not true duplicates. However, given the rapid advancements in large language models since this analysis was run, we suspect that more recent iterations of this technology may yield significantly improved results. Table A3 in the supplementary materials provides an example of the most highly similar item pairs identified by the model, along with the human ratings.

3.5. Step 5: Pilot testing

Participants were enrolled between August 2023 and February 2024 across eight primary care clinics in Bhopal. A total of 200 individuals participated in testing the baseline battery, including 125 individuals with moderate to severe depressive symptoms (PHQ-9 score $\geq$ 10) and 75 individuals with PHQ-9 scores <10. The characteristics of the participants are summarized in Table A4 in the supplementary materials. Participants had a mean age of 38 years (SD = 14.6), 47% identified as female, and the majority were married. Educational attainment averaged 10 years of schooling. Self-reported Hindi reading ability varied across participants: approximately 10% reported poor or very poor reading ability, and a further 18% rated their ability as acceptable, suggesting that over a quarter of the sample had limited literacy. Half of the participants were employed in diverse occupations, and 35% were unemployed. Participants were distributed across PHQ-9 depressive severity categories, with one-third reporting minimal or mild symptoms, 31% moderate symptoms, 21% moderately severe symptoms, and 11% severe.

3.5.1. Findings from qualitative data

We analyzed two sources of qualitative data: 514 structured feedback surveys completed by interviewers and 40 audio-recorded baseline assessments coded by a research assistant. Both sources highlighted the

Table 2
Baseline constructs and measurement tools.

Domain	Construct	Measure	# Items	Brevity	Prior validation	Language availability	Public accessibility	# criteria
Sociodemographic	Age	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Sex	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Gender	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Marital status	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Children	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Education	2-items	2	Yes	Yes	Yes	Yes	4
Sociodemographic	Literacy skills	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Employment	1-item	2	Yes	Yes	Yes	Yes	4
Sociodemographic	Income level	1-item	3	Yes	Yes	Yes	Yes	4
Sociodemographic	Caste	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Religion	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Urban/rural setting	1-item	1	Yes	Yes	Yes	Yes	4
Sociodemographic	Poverty status	Simple Poverty Scorecard India	10	No	Yes	No	Yes	2
Clinical	Depression severity	Patient Health Questionnaire (PHQ-9)	9	Yes	Yes	Yes	Yes	4
Clinical	Age of onset	1-item	1	Yes	Yes	No	Yes	3
Clinical	Chronicity	1-item	1	Yes	Yes	No	Yes	3
Clinical	Recurrent episodes	2-items	2	Yes	Yes	No	Yes	3
Clinical	Past psychiatric treatment	3-items	3	Yes	Yes	No	Yes	3
Clinical	Medication adherence	3-items	3	Yes	No	No	Yes	2
Clinical	Depression subtype	Inventory of Depressive Symptomatology – Self Report (IDS-SR)	8	Yes	Yes	No	Yes	3
Clinical	Somatic symptoms	Patient Health Questionnaire Somatic Symptom Scale (PHQ-15)	15	No	Yes	No	Yes	2
Clinical	Side effects	Antidepressant Side-Effect Checklist (ASEC)	16	No	Yes	No	Yes	2
Clinical	Hopelessness	Beck Hopelessness Scale – Short Version (BHS-4)	4	Yes	Yes	No	Yes	3
Clinical	Patient treatment preferences	1-item	1	Yes	No	No	Yes	2
Clinical	Family treatment preferences	1-item	1	Yes	No	No	Yes	2
Clinical	Outcome expectations	1-item	1	Yes	No	No	Yes	2
Clinical	Positive/negative affect	Positive and Negative Affect Schedule – Short Form (PANAS-SF)	10	No	No	No	Yes	1
Clinical	Worry	Penn State Worry Questionnaire – Abbreviated (PSWQ-A)	8	Yes	Yes	No	Yes	3
Clinical	Well-being	World Health Organization Well-Being Index (WHO-5)	5	Yes	Yes	No	Yes	3
Clinical	Physical illness	Composite International Diagnostic Interview (CIDI) - physical illness items	3	Yes	Yes	No	Yes	3
Clinical	Obesity	Body Mass Index (BMI) based on weight/height	2	Yes	Yes	N.A	Yes	3
Clinical	Anhedonia	Snaith-Hamilton Pleasure Scale (SHAPS)	14	No	Yes	No	Yes	2
Clinical	Patient activation	PREMIUM Abbreviated Activation Scale (PAAS)	5	Yes	Yes	Yes	Yes	4
Clinical	Rumination	Analytical Rumination Questionnaire (ARQ)	9	Yes	No	No	Yes	2
Psychiatric comorbidity	Suicidality	Columbia Suicide Severity Rating Scale – Screener (C-SSRS)	7	Yes	Yes	Yes	Yes	4
Psychiatric comorbidity	Anxiety disorder	Generalized Anxiety Disorder Scale (GAD-7)	7	Yes	Yes	Yes	Yes	4
Psychiatric comorbidity	Alcohol use	Alcohol Use Disorders Identification Test – Concise (AUDIT-C)	3	Yes	Yes	No	Yes	3
Psychiatric comorbidity	Smoking	1-item	5	Yes	Yes	Yes	Yes	4
Psychiatric comorbidity	Personality disorder	Standardized Assessment of Personality – Abbreviated Scale (SAPAS)	8	Yes	Yes	No	Yes	3
Psychiatric comorbidity	Sleep problems	Brief Pittsburgh Sleep Quality Index (B-PSQI)	6	Yes	Yes	No	No	2
Psychiatric comorbidity	Alexithymia	Toronto Alexithymia Scale (TAS-20)	20	No	Yes	Yes	Yes	3
Psychiatric comorbidity	Panic disorder	Panic Disorder Severity Scale (PDSS)	2	Yes	Yes	No	Yes	3

(*continued on next page*)

Table 2 (*continued*)

Domain	Construct	Measure	# Items	Brevity	Prior validation	Language availability	Public accessibility	# criteria
Psychiatric comorbidity	Trauma	Post-Traumatic Stress Disorder Checklist (PCL)	6	Yes	Yes	No	Yes	3
Personality	Attachment style	Experiences in Close Relationships Scale – Short Form (ECR-12)	12	No	Yes	No	Yes	2
Personality	Neuroticism & extraversion	Big Five Inventory – Short Form (BFI-10)	10	No	Yes	No	Yes	2
Personality	Anxiety sensitivity	Short Scale Anxiety Sensitivity Index (SSASI)	5	Yes	Yes	No	Yes	3
Personality	Self-efficacy	General Self-Efficacy Scale (GSE)	10	No	Yes	Yes	Yes	3
Personality	Self-esteem	Rosenberg Self-Esteem Scale (RSES)	10	No	Yes	No	Yes	2
Personality	Resilience	Abbreviated Connor-Davidson Resilience Scale (CD-RISC-2)	2	Yes	Yes	No	No	2
Personality	Dysfunctional attitudes	Dysfunctional Attitude Scale (DAS-A-17)	17	No	Yes	No	Yes	2
Personality	Emotion regulation	Emotion Regulation Questionnaire (ERQ)	10	No	Yes	Yes	Yes	3
Life stress	Childhood abuse	Adverse Childhood Experiences Questionnaire (ACE)	10	No	Yes	Yes	Yes	3
Life stress	High stress	Perceived Stress Scale (PSS-10)	10	No	Yes	No	Yes	2
Life stress	Intimate partner violence	2-items	2	Yes	Yes	Yes	Yes	4
Life stress	Loneliness	UCLA Loneliness Scale (UCLA-3)	3	Yes	Yes	No	Yes	3
Life stress	Grief	Prolonged Grief Disorder items + Brief Grief Questionnaire	4	Yes	No	No	Yes	2
Life stress	Social support	Multidimensional Scale of Perceived Social Support (MSPSS)	12	No	Yes	Yes	Yes	3
Impact	Disability	World Health Organization Disability Assessment Schedule II (WHODAS-II)	12	No	Yes	Yes	Yes	3
Impact	Physical activity	International Physical Activity Questionnaire – Short Form (IPAQ-SF)	7	Yes	Yes	No	Yes	3
Impact	Social functioning	Social Functioning Questionnaire (SFQ)	8	Yes	Yes	No	Yes	3
Neurocognitive	Working memory	Digit Span Task	–	N.A	Yes	Yes	Yes	3
Neurocognitive	Working memory	Constructional Praxis Task (delayed recall)	–	N.A	Yes	Yes	Yes	3
Neurocognitive	Inhibition	Symbol Cancellation Task	–	N.A	Yes	Yes	Yes	3
Neurocognitive	Shifting	Trail Making Task	–	N.A	Yes	Yes	Yes	3
Neurocognitive	Reinforcement learning	Category learning task	–	N.A	Yes	No	No	1
Neurocognitive	Intellectual functioning	RAVEN's task	–	N.A	Yes	Yes	Yes	3
Neurocognitive	Intellectual functioning	Serial 7 s task	–	N.A	Yes	Yes	Yes	3
Genetic risk	Polygenic Risk Scores (PRS) based on a genome wide microarray based genotyping (GSA – Global Screening Array)	Depression and psychiatric traits	–	N.A	N.A	N.A	N.A	N.A
Biological	Inflammation	C-Reactive Protein (CRP)	–	N.A	N.A	N.A	N.A	N.A
Biological	Pharmacokinetics	Cytochrome P450 2D6 genotype (CYP2D6)	–	N.A	N.A	N.A	N.A	N.A

excessive length of the baseline assessment as a key concern. Participants, particularly older adults, individuals with limited literacy, and those with visual impairments, found the survey burdensome and often lost interest before completion. Many struggled with structured response formats, particularly Likert scales, as they preferred to describe their experiences in their own words rather than choose from predefined options. As a result, interviewers frequently had to repeat or rephrase items, further increasing administration time. Several measures proved particularly challenging to administer. Complex scales, such as those assessing depression subtype (Inventory of Depressive Symptomatology, IDS-SR), rumination (Analytical Rumination Questionnaire, ARQ), emotional regulation (Emotion Regulation Questionnaire, ERQ), and dysfunctional attitudes (Dysfunctional Attitudes Scale, DAS), were frequently misread by interviewers or required extensive explanation. Participants commonly sought clarification for items assessing anhedonia (Snaith-Hamilton Pleasure Scale, SHAPS), somatic symptoms (Patient Health Questionnaire, PHQ-15), personality traits (BFI), and alexithymia (TAS-20). These findings prompted additional interviewer training, development of supportive materials, and, in some cases, simplification of item wording (e.g., IDS-SR). Additionally, one of the cognitive tasks, the reinforcement learning task, was frequently described by both interviewers and participants as overly long and difficult for participants to understand or meaningfully engage with. To reduce the burden, we added a practice run and the option to stop after 40 trials (the minimum needed for analysis). Sensitive topics posed additional challenges. Questions on suicide severity (Columbia Suicide Severity Rating Scale, C-SSRS), childhood trauma (Adverse Childhood Experiences, ACE), and intimate partner violence often made participants visibly uncomfortable, occasionally leading to distress. Relatedly, substance use and sexual function questions were particularly sensitive

for female participants. To mitigate harm, interviewers were trained to handle sensitive topics and support distress; participants could skip questions or stop anytime, and all received a support resource list. Beyond comprehension and sensitivity issues, logistical and technical barriers also impacted data collection. Interviewers reported frequent environmental disruptions, including construction noise at clinics and interruptions from accompanying family members, which compromised privacy and may have influenced responses. Additionally, intermittent tablet malfunctions and server failures impeded smooth administration. In response, we revised site protocols to improve privacy, enhanced on-site technical support, and prioritized offline data collection software (i.e., REDCap).

3.5.2. Findings from quantitative data

The quantitative analysis supported and expanded upon several themes that emerged from the qualitative data, particularly regarding the overall length of the assessment, as well as the limited variability observed in responses to certain measures. The baseline assessment took an average of 3 h and 7 min to complete (95% CI: 2 h 53 min – 3 h 21 min), including two planned breaks. However, in approximately one-third of cases, the interview had to be split across multiple days due to participant fatigue or limited availability, reinforcing concerns raised in the qualitative analysis. The longest scales, such as those assessing depression subtype (IDS-SR), alexithymia (TAS-20), and anhedonia (SHAPS), were also those frequently cited as problematic by interviewers and participants (see Fig. A3 in the supplementary materials). The majority of clinical and psychological constructs showed adequate distributional properties with no floor or ceiling effects (Supplementary Table A5). Exceptions were most notable for the substance use measure (ASSIST), where fewer than 3% of participants reported ever using stimulants, sedatives, or opioids, indicating that these questions may not capture meaningful variation in this population; contextually relevant adaptations were therefore made, such as adding smokeless tobacco products (e.g., chewing) given their prevalence in India (Rani et al., 2003). Floor effects were also observed for some low-prevalence clinical experiences, including panic disorder (68.1%) and grief (51.1%). Pairwise Spearman correlations across all battery scales are presented in Supplementary Fig. A4. The heatmap indicates broad dimensional coverage with limited redundancy across domains, though strong correlations were observed within the clinical symptom cluster (particularly among anxiety, PTSD, worry, negative affect, and somatic symptoms, $r = 0.50–0.73$). Of the 76 participants approached for blood sampling, 51 consented and 25 declined, most commonly due to fear of needles (52%) or general apprehension about blood draws (20%). These findings led to the inclusion of a saliva collection option in the main trial. With the addition of saliva sampling as a less invasive alternative, 97% of participants in the main trial consented to provide either a blood or saliva sample. Although some biomarkers (e.g., C-reactive protein) cannot be assessed using saliva, this approach substantially improved participation in biological data collection.

3.6. Step 6: Item reduction analyses

Most scales exhibited a single-factor structure with strong item loadings (>0.4). However, a few measures indicated a two-factor structure, including the Positive and Negative Affect Schedule (PANAS-SF), which separated into positive and negative affect dimensions, and the Emotion Regulation Questionnaire (ERQ), which split into reappraisal and suppression factors. For certain constructs, the factor structure did not align with the original scale (e.g., ARQ, SHAPS, ERQ), warranting further refinement. Additionally, some scales, such as the Toronto Alexithymia Scale (TAS-20), contained items with weak factor loadings, reinforcing the need for item reduction. We then applied Lasso regression to identify the most predictive items within each construct, reducing redundancy while preserving construct coverage. This resulted in substantial item reductions for certain scales, particularly for longer measures, such as Alexithymia (TAS-20, reduced from 20 to 5 items) and Dysfunctional Attitudes (DAS-17, reduced from 17 to 5 items). However, some constructs, such as Childhood Trauma (ACE) and Depression Subtype (IDS-SR), were retained in full as they are well-supported in the literature as potential prescriptive indicators. In these cases, the study team prioritized complete item coverage to preserve clinically meaningful information that might be lost through reduction. Overall, the combination of both analyses led to a reduction in total items from 389 to 284, while preserving all core constructs identified in previous steps. This refinement significantly shortened the baseline assessment, reducing administration time from over 3 h to approximately 2 h, improving feasibility for both participants and research staff. A full summary of these results, item reductions, and retained items is presented in Table A6 in the supplementary materials. The full list of items included in the OptimizeD baseline battery is provided in Table A7.

4. Discussion

This study outlines a systematic, stepwise multimethod approach to developing a baseline assessment battery for a precision treatment trial aimed at predicting differential response to two first-line treatments for depression: behavioral activation and fluoxetine. Drawing on findings from a previous review, expert input, pilot testing, and advanced statistical techniques, we sought to balance the breadth of relevant predictors with the brevity and feasibility required for implementation in real-world primary care settings. The resulting battery retains the full range of constructs identified in prior literature and expert consultation, while reducing administration time to approximately two hours. We acknowledge that this remains substantial and is not intended for routine clinical use at this stage. However, we anticipate that further reductions in length and complexity will be possible following the main trial, since not all measures will contribute to the resulting algorithm, improving the battery's scalability for use in routine depression care.

The final battery reflects a deliberately broad, multidimensional framework designed to capture the complex and heterogeneous nature of depression and its treatment response. It includes 68 constructs across clinical, psychological, cognitive, social, and biological domains – balancing well-established predictors (e.g., depression severity, comorbid anxiety) with less commonly assessed yet potentially important factors (e.g., personality traits). This approach addresses a key limitation of much prior research, which has often relied on predictors drawn from a single domain, limiting the ability to capture the complexity of treatment response. Importantly, the battery also integrates multiple modes of assessment – including self-report questionnaires, neurocognitive tasks, and biological markers – each contributing complementary insights. While questionnaires capture subjective symptoms and psychosocial context, neurocognitive tasks provide objective insight into core cognitive functions such as memory, attention, and executive control. Biological measures complement these by offering data on genetic vulnerability through the generation of a polygenic risk score based on genome-wide microarray-based genotyping, assessment of the drug metabolising enzyme gene CYP2D6 genotype in the context of fluoxetine response, and underlying inflammatory processes. This multimodal design allows for a more comprehensive assessment of individual differences, which may enhance prediction accuracy and support the development of more mechanistically informed treatment models. Although biological markers are less scalable in primary care contexts (though the availability of these is expanding in primary care), one of the exploratory aims of the OptimizeD trial is to examine whether their inclusion improves predictive performance when added to the clinical decision tool (Pozuelo et al., 2025). These findings will help clarify when biological data enhances clinical utility.

Although several constructs in the battery were originally developed in Western contexts, instrument selection prioritized measures previously validated in India or similar low-resource settings. The retained

constructs were supported by evidence from both the systematic review and expert consultation, suggesting their theoretical relevance for differential treatment response. However, whether these constructs carry the same conceptual meaning in rural Indian populations – and whether they function as effective treatment moderators in this context – remains an open question. Qualitative feedback from interviewers, for example, suggested that items relating to emotional awareness and dysfunctional cognitions sometimes required substantial contextual explanation. The OptimizeD trial will provide the first empirical test of whether these constructs generalize as treatment moderators in an Indian primary care population.

As part of the development process, we conducted a semantic similarity analysis to identify potential item-level redundancy across scales. While promising in principle, the model proved unreliable in practice and often failed to distinguish conceptually distinct items. In addition, the analysis was designed to detect redundancy at the item level rather than at the construct level, and could have therefore missed overlap between scales measuring the same underlying construct (although Fig. A4 suggests limited construct-level overlap). Future applications would benefit from more rigorous preprocessing to remove shared syntactic scaffolding, combining item-level with construct-level analyses, and using more advanced embedding models. More broadly, these approaches are best used to flag potential redundancies for expert review rather than as standalone decision-making tools, although this may change as the technology continues to advance rapidly.

Beyond the empirical results, this work contributes a practical example of applied, data-driven instrument design. It stands in contrast to common practices in clinical research where measures are selected based on precedent or investigator preference. Our structured, evidence-based process strengthens internal validity, replicability, and generalizability, especially in applied and cross-cultural settings. In addition, the resulting battery may serve as a practical template for researchers designing baseline assessments for precision treatment trials in depression and related areas. While grounded in depression research, the principles and framework we describe are broadly relevant for researchers designing multidimensional assessments under real-world constraints.

Despite its strengths, this study has several limitations. First, although the selected constructs were drawn from a global evidence base, the battery was designed specifically for use in Hindi in primary care settings in India. However, the grounding of the battery in international literature enhances its potential adaptability and relevance across diverse contexts globally. Second, most predictors were measured using questionnaires, which are subject to recall bias, social desirability bias, and comprehension difficulties. While we selected validated, context-appropriate tools to minimize these issues, the inclusion of neurocognitive tasks and biological measures in the battery may also help address these limitations. Third, the systematic review update relied solely on PubMed, following the methodology of the original review. While PubMed captures much of the medical literature, psychotherapy-focused studies may be better indexed in psychology-specific databases (e.g., PsycINFO), potentially resulting in under-identification of certain psychotherapy-specific predictors. This limitation was partially mitigated by expert consultation, which included authors of key predictive and treatment studies, to identify constructs that may have been overlooked. Fourth, the expert consultation sample was derived largely from authors identified through the systematic review, and 40% of respondents were authors of included studies. Although this strategy ensured domain expertise, it may have introduced confirmation bias if experts overestimated the importance of constructs they had previously investigated. However, inclusion in the final precision treatment rule depended on the construct's empirical performance as a predictor of response (and differential response), rather than on expert ratings. In addition, most respondents were affiliated with institutions in high-income countries, although more than half had prior research experience in low- and middle-income settings. This composition may nevertheless limit the extent to which the consultation reflects locally embedded perspectives. The survey also used a binary relevance rating ("likely" vs. "unlikely"), which simplified responses but may have limited the nuance with which experts could express their views. Fifth, efforts to reduce participant burden led to the removal or modification of some items, which may have weakened construct validity or excluded potentially informative predictors. In addition, the item reduction analyses were conducted in a relatively small analytic subsample, which may limit the stability and generalisability of the selected items. Sixth, individual indicators most strongly associated with a composite measure need not necessarily be the most predictive of outcomes (VanderWeele, 2022). The OptimizeD trial will provide an opportunity to assess whether these trade-offs affected the ability to detect moderation effects. If no moderation effects are detected, it will be challenging to determine whether this is due to a theoretical failure (i.e., no meaningful differential treatment response exists) or a measurement failure (i.e., the baseline battery did not capture predictors effectively). We will address both these possibilities by comparing the primary care sample recruited for this study to secondary care samples from prior trials to assess differences in key variables such as severity, chronicity, and prior treatment history (DeRubeis et al., 2005) and by comparing the psychometric properties of the refined measures in the full trial with those of the battery tested in the pilot to assess whether item reduction compromised measurement precision.

A key strength of this work is that the development of the battery was grounded in empirical evidence, expert consensus, and feasibility constraints relevant to low-resource healthcare settings. In doing so, this study moves beyond theoretical models of personalization toward the creation of a practical, scalable tool that can be integrated into primary care settings and inform real-time clinical decision-making. By identifying the most effective treatment for each patient at baseline, the approach holds potential to improve patient outcomes while reducing inefficiencies and strain on limited mental health resources. As digital and AI-driven decision-support systems continue to evolve, structured and streamlined baseline assessments, such as this one, may serve as foundations for automated triage tools, increasing reach and cost-effectiveness at scale.

In summary, this study presents a structured and replicable framework for designing scalable, data-driven baseline assessments to support precision mental health in low-resource settings. The current battery was designed to be comprehensive, serving both as a foundation for the OptimizeD trial and as a potential template for researchers conducting precision trials for depression management. OptimizeD will identify which variables (or combinations of variables) best predict optimal treatment assignment, providing proof of concept for implementing personalized depression care in primary care. More broadly, this approach can be adapted for other mental health conditions and integrated into clinical and policy frameworks to improve treatment efficiency and equity. As the field moves toward more individualized care, practical tools like this will be critical to closing the treatment gap and delivering better outcomes for diverse patient populations.

CRediT authorship contribution statement

Julia R. Pozuelo: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Ronald C. Kessler:** Writing – review & editing, Funding acquisition, Formal analysis, Conceptualization. **Vikram Patel:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Anuja Lahiri:** Writing – review & editing, Project administration, Data curation. **Yashika Parashar:** Writing – review & editing, Project administration, Data curation. **Mohammad M. Herzallah:** Writing – review & editing, Funding acquisition, Data curation. **Mimansa Khanduri:** Writing – review & editing, Project administration, Data curation. **Kathryn O'Neill:** Writing – review & editing, Data curation. **Alyssa Martinez:** Writing – review & editing, Data curation. **Anant Bhan:**

Writing – review & editing, Funding acquisition. **Daisy R. Singla:** Writing – review & editing, Funding acquisition. **John A. Naslund:** Writing – review & editing, Data curation. **Pim Cuijpers:** Writing – review & editing, Funding acquisition. **Jordan W. Smoller:** Writing – review & editing, Funding acquisition. **Karmel W. Choi:** Writing – review & editing, Funding acquisition. **Tyler J. VanderWeele:** Writing – review & editing, Funding acquisition. **Rahul S.P. Singh:** Writing – review & editing, Project administration, Data curation. **Akanksha Shukla:** Writing – review & editing. **Arvind Kushwah:** Writing – review & editing, Project administration. **Azaz Khan:** Writing – review & editing. **Sruthi G.:** Writing – review & editing. **Varun Shende:** Writing – review & editing. **Yashwant K. Mehra:** Writing – review & editing. **Michelle Melwyn Joel:** Writing – review & editing. **Shubham Atal:** Writing – review & editing. **Abhijit Rozatkar:** Writing – review & editing, Funding acquisition. **Tamonud Modak:** Writing – review & editing. **Steven D. Hollon:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Ethical standards

All study procedures involving human subjects received ethics approval from institutional review boards in India (Sangath IRB: AB-2021-69; AIIMS Bhopal: EF0237) and the United States (Harvard Medical School IRB20-2144). The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2024.

Funding statement

This work was supported by the National Institute of Mental Health (NIMH, grant number: R01MH121632-01).

Declaration of competing interest

In the past 3 years, Dr. Kessler was a consultant for Cambridge Health Alliance, Child Mind Institute, Massachusetts General Hospital, RallyPoint LLC., Sage Therapeutics, University of Michigan, and University of North Carolina. He has stock options in Cerebral Inc., Mirah, PYM (Prepare Your Mind), and Verisense Health. He has an ownership interest in Menssano LLC.

Acknowledgments

We thank all the participants who generously took part in the pilot study, as well as the staff working in primary healthcare facilities who facilitated their recruitment. We are also grateful to the experts who took the time to complete the survey and share their insights. Finally, we thank the research staff whose dedication enables the daily implementation of the OptimizeD trial.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2026.121674.

References

Akechi, T., Mantani, A., Kurata, K., Hirota, S., Shimodera, S., Yamada, M., Inagaki, M., Watanabe, N., Kato, T., Furukawa, T.A., 2019. Predicting relapse in major depression after successful initial pharmacological treatment. J. Affect. Disord. 250, 108–113. https://doi.org/10.1016/j.jad.2019.03.004.

Allen, T.A., Harkness, K.L., Lam, R.W., Milev, R., Frey, B.N., Mueller, D.J., Uher, R., Kennedy, S.H., Quilty, L.C., 2021. Interactions between neuroticism and stressful life events predict response to pharmacotherapy for major depression: a CAN-BIND 1 report. Personal. Ment. Health 15, 273–282. https://doi.org/10.1002/pmh.1514.

Arnow, B.A., Blasey, C., Williams, L.M., Palmer, D.M., Rekshan, W., Schatzberg, A.F., Etkin, A., Kulkarni, J., Luther, J.F., Rush, A.J., 2015. Depression subtypes in predicting antidepressant response: a report from the iSPOT-D trial. AJP 172, 743–750. https://doi.org/10.1176/appi.ajp.2015.14020181.

Bauer, M., Thase, M., Liu, S., Earley, W., Eriksson, H., 2015. Analysis of potentially predictive factors of efficacy of adjunct extended-release quetiapine fumarate in patients with major depressive disorder. J. Psychopharmacol. 29, 565–574. https://doi.org/10.1177/0269881114552715.

Bausch, P., Fangmeier, T., Zobel, I., Schoepf, D., Drost, S., Schnell, K., Walter, H., Berger, M., Normann, C., Schramm, E., 2017. The impact of childhood maltreatment on the differential efficacy of CBASP versus escitalopram in patients with chronic depression: a secondary analysis. Clin. Psychol. Psychother. 24, 1155–1162. https://doi.org/10.1002/cpp.2081.

Békés, V., Dunkley, D.M., Taylor, G., Zuroff, D.C., Lewkowski, M., Elizabeth Foley, J., Myhr, G., Westreich, R., 2015. Chronic stress and attenuated improvement in depression over 1 year: the moderating role of perfectionism. Behav. Ther. 46, 478–492. https://doi.org/10.1016/j.beth.2015.02.003.

Bucher, M.A., Suzuki, T., Samuel, D.B., 2019. A meta-analytic review of personality traits and their associations with mental health treatment outcomes. Clin. Psychol. Rev. 70, 51–63. https://doi.org/10.1016/j.cpr.2019.04.002.

Buckman, J.E.J., Saunders, R., Fearon, P., Leibowitz, J., Pilling, S., 2019. Attentional control as a predictor of response to psychological treatment for depression and relapse up to 1 year after treatment: a pilot cohort study. Behav. Cogn. Psychother. 47, 318–331. https://doi.org/10.1017/S1352465818000590.

Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., Corlett, P.R., 2016. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 3, 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X.

Chekroud, A.M., Gueorguieva, R., Krumholz, H.M., Trivedi, M.H., Krystal, J.H., McCarthy, G., 2017. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. JAMA Psychiatry 74, 370–378. https://doi.org/10.1001/jamapsychiatry.2017.0025.

Chekroud, A.M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., Choi, K., 2021. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 20, 154–170. https://doi.org/10.1002/wps.20882.

Chui, H., Zilcha-Mano, S., Dinger, U., Barrett, M.S., Barber, J.P., 2016. Dependency and self-criticism in treatments for depression. J. Couns. Psychol. 63, 452–459. https://doi.org/10.1037/cou0000142.

Cohen, Z.D., DeRubeis, R.J., 2018. Treatment selection in depression. Annu. Rev. Clin. Psychol. 14, 209–236. https://doi.org/10.1146/annurev-clinpsy-050817-084746.

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

Constantinou, M.P., Frueh, B.C., Fowler, J.C., Allen, J.G., Madan, A., Oldham, J.M., Fonagy, P., 2020. Predicting depression outcomes throughout inpatient treatment using the general and specific personality disorder factors. Psychol. Med. 1–9. https://doi.org/10.1017/S003329172000361X.

Cuijpers, P., Reynolds, C.F., Donker, T., Li, J., Andersson, G., Beekman, A., 2012. Personalized treatment of adult depression: medication, psychotherapy, or both? A systematic review. Depress. Anxiety 29, 855–864. https://doi.org/10.1002/da.21985.

Cuijpers, P., Sijbrandij, M., Koole, S.L., Andersson, G., Beekman, A.T., Reynolds, C.F., 2013. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry 12, 137–148. https://doi.org/10.1002/wps.20038.

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

Cuijpers, P., Miguel, C., Ciharova, M., Harrer, M., Moir, F., Roskvist, R., van Straten, A., Karyotaki, E., Arroll, B., 2023. Psychological treatment of adult depression in primary care compared with outpatient mental health care: a meta-analysis. J. Affect. Disord. 339, 660–675. https://doi.org/10.1016/j.jad.2023.07.011.

Dawson, E.L., Caveney, A.F., Meyers, K.K., Weisenbach, S.L., Giordani, B., Avery, E.T., Schallmo, M.P., Bahadori, A., Bieliauskas, L.A., Mordhorst, M., Marcus, S.M., Kerber, K., Zubieta, J.K., Langenecker, S.A., 2017. Executive functioning at baseline prospectively predicts depression treatment response. Prim Care Companion CNS Disord 19. https://doi.org/10.4088/PCC.16m01949.

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.

DeRubeis, R.J., Hollon, S.D., Amsterdam, J.D., Shelton, R.C., Young, P.R., Salomon, R.M., O'Reardon, J.P., Lovett, M.L., Gladis, M.M., Brown, L.L., Gallop, R., 2005. Cognitive therapy vs medications in the treatment of moderate to severe depression. Arch. Gen. Psychiatry 62, 409–416. https://doi.org/10.1001/archpsyc.62.4.409.

DeRubeis, R.J., Cohen, Z.D., Forand, N.R., Fournier, J.C., Gelfand, L.A., Lorenzo-Luaces, L., 2014. The personalized advantage index: translating research on prediction into individualized treatment recommendations. A demonstration. PLoS One 9, e83875. https://doi.org/10.1371/journal.pone.0083875.

Dinger, U., Ehrenthal, J.C., Nikendei, C., Schauenburg, H., 2017. Change in self-esteem predicts depressive symptoms at follow-up after intensive multimodal psychotherapy for major depression. Clin. Psychol. Psychother. 24, 1040–1046. https://doi.org/10.1002/cpp.2067.

Driessen, E., Cuijpers, P., Hollon, S.D., Dekker, J.J.M., 2010. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. J. Consult. Clin. Psychol. 78, 668–680. https://doi.org/10.1037/a0020570.

Dunlop, K., Rizvi, S.J., Kennedy, S.H., Hassel, S., Strother, S.C., Harris, J.K., Zamyadi, M., Arnott, S.R., Davis, A.D., Mansouri, F., Schulze, L., Ceniti, A.K., Lam, R.W., Milev, R., Rotzinger, S., Foster, J.A., Frey, B.N., Parikh, S.V., Soares, C.N., Uher, R., Turecki, G., MacQueen, G.M., Downar, J., 2020. Clinical, behavioral, and neural measures of reward processing correlate with escitalopram response in depression: a Canadian biomarker integration network in depression (CAN-BIND-1) report. Neuropsychopharmacol 45, 1390–1397. https://doi.org/10.1038/s41386-020-0688-x.

Edmonds, M., Hadjistavropoulos, H.D., Schneider, L.H., Dear, B.F., Titov, N., 2018. Who benefits most from therapist-assisted internet-delivered cognitive behaviour therapy in clinical practice? Predictors of symptom change and dropout. J. Anxiety Disord. 54, 24–32. https://doi.org/10.1016/j.janxdis.2018.01.003.

Ermers, N.J., Hagoort, K., Scheepers, F.E., 2020. The predictive validity of machine learning models in the classification and treatment of major depressive disorder: state of the art and future directions. Front. Psychol. 11. https://doi.org/10.3389/fpsyt.2020.00472.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., Strahan, E.J., 1999. Evaluating the use of exploratory factor analysis in psychological research. Psychol. Methods 4, 272–299. https://doi.org/10.1037/1082-989X.4.3.272.

Ferrari, A.J., Santomauro, D.F., Aali, A., Abate, Y.H., Abbafati, C., Abbastabar, H., Abd ElHafeez, S., Abdelmasseh, M., Abd-Elsalam, S., Abdollahi, A., others, 2024. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. Lancet 403, 2133–2161.

Fitzpatrick, O.M., Whelen, M.L., Falkenström, F., Strunk, D.R., 2020. Who benefits the most from cognitive change in cognitive therapy of depression? A study of interpersonal factors. J. Consult. Clin. Psychol. 88, 128–136. https://doi.org/10.1037/ccp0000463.

Fonseka, T.M., MacQueen, G.M., Kennedy, S.H., 2018. Neuroimaging biomarkers as predictors of treatment outcome in major depressive disorder. J. Affect. Disord. 233, 21–35. https://doi.org/10.1016/j.jad.2017.10.049.

Forand, N.R., DeRubeis, R.J., 2014. Extreme response style and symptom return after depression treatment: the role of positive extreme responding. J. Consult. Clin. Psychol. 82, 500–509. https://doi.org/10.1037/a0035755.

Fournier, J.C., DeRubeis, R.J., Shelton, R.C., Hollon, S.D., Amsterdam, J.D., Gallop, R., 2009. Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. J. Consult. Clin. Psychol. 77, 775–787. https://doi.org/10.1037/a0015401.

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R. C., Fawcett, J., 2010. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA 303, 47–53. https://doi.org/10.1001/jama.2009.1943.

Fowler, J.C., Patriquin, M., Madan, A., Allen, J.G., Frueh, B.C., Oldham, J.M., 2015. Early identification of treatment non-response utilizing the patient health questionnaire (PHQ-9). J. Psychiatr. Res. 68, 114–119. https://doi.org/10.1016/j.jpsychires.2015.06.018.

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

Gaspersz, R., Lamers, F., Kent, J.M., Beekman, A.T.F., Smit, J.H., van Hemert, A.M., Schoevers, R.A., Penninx, B.W.J.H., 2017. Anxious distress predicts subsequent treatment outcome and side effects in depressed patients starting antidepressant treatment. J. Psychiatr. Res. 84, 41–48. https://doi.org/10.1016/j.jpsychires.2016.09.018.

Gerra, M.L., Marchesi, C., Amat, J.A., Blier, P., Hellerstein, D.J., Stewart, J.W., 2014. Does negative affectivity predict differential response to an SSRI versus a non-SSRI antidepressant? J. Clin. Psychiatry 75 (9), e939–e944. https://doi.org/10.4088/JCP.14m09025.

Gollan, J.K., Hoxha, D., Hunnicutt-Ferguson, K., Norris, C.J., Rosebrock, L., Sankin, L., Cacioppo, J., 2016. The negativity bias predicts response rate to behavioral activation for depression. J. Behav. Ther. Exp. Psychiatry 52, 171–178. https://doi.org/10.1016/j.jbtep.2015.09.011.

Gov. of India, 2011. Census 2011.

Hallgren, M., Nakitanda, O.A., Ekblom, Ö., Herring, M.P., Owen, N., Dunstan, D., Helgadottir, B., Forsell, Y., 2016. Habitual physical activity levels predict treatment outcomes in depressed adults: a prospective cohort study. Prev. Med. 88, 53–58. https://doi.org/10.1016/j.ypmed.2016.03.021.

Hallgren, M., Lundin, A., Tee, F.Y., Burström, B., Forsell, Y., 2017. Somebody to lean on: social relationships predict post-treatment depression severity in adults. Psychiatry Res. 249, 261–267. https://doi.org/10.1016/j.psychres.2016.12.060.

Hornstein, S., Forman-Hoffman, V., Nazander, A., Ranta, K., Hilbert, K., 2021. Predicting therapy outcome in a digital mental health intervention for depression and anxiety: a machine learning approach. Digit Health 7, 20552076211060659. https://doi.org/10.1177/20552076211060659.

Iyer, G.K., Paplikar, A., Alladi, S., Dutt, A., Sharma, M., Mekala, S., Kaul, S., Saroja, A.O., Divyaraj, G., Ellajosyula, R., Ghosh, A., Hooda, R., Justus, S., Kandukuri, R., Khan, A. B., Mathew, R., Mathuranath, P.S., Menon, R., Nandi, R., Narayanan, J., Nehra, A., Padma, M.V., Pauranik, A., Ramakrishnan, S., Sabnis, P., Sarath, L., Shah, U., Tripathi, M., Sylaja, P.N., Varma, R.P., Verma, M., Varghese, F., Consortium, I.N.T.B, 2020. Standardising dementia diagnosis across linguistic and educational diversity: study design of the Indian Council of medical research-neurocognitive tool box (ICMR-NCTB). J. Int. Neuropsychol. Soc. 26, 172–186. https://doi.org/10.1017/S1355617719001127.

Jha, M.K., Minhajuddin, A., Greer, T.L., Carmody, T., Rush, A.J., Trivedi, M.H., 2016. Early improvement in work productivity predicts future clinical course in depressed outpatients: findings from the CO-MED trial. Am. J. Psychiatry 173, 1196–1204. https://doi.org/10.1176/appi.ajp.2016.16020176.

Jha, M.K., Malchow, A.L., Grannemann, B.D., Rush, A.J., Trivedi, M.H., 2018. Do baseline sub-threshold hypomanic symptoms affect acute-phase antidepressant outcome in outpatients with major depressive disorder? Preliminary findings from the randomized CO-MED trial. Neuropsychopharmacology 43, 2197–2203. https://doi.org/10.1038/s41386-018-0180-z.

Kaiser, T., Volkmann, C., Volkmann, A., Karyotaki, E., Cuijpers, P., Brakemeier, E.-L., 2022. Heterogeneity of treatment effects in trials on psychotherapy of depression. Clin. Psychol. Sci. Pract. 29, 294–303. https://doi.org/10.1037/cps0000079.

Kaneriya, S.H., Robbins-Welty, G.A., Smagula, S.F., Karp, J.F., Butters, M.A., Lenze, E.J., Mulsant, B.H., Blumberger, D., Anderson, S.J., Dew, M.A., Lotrich, F., Aizenstein, H. J., Diniz, B.S., Reynolds III, C.F., 2016. Predictors and moderators of remission with aripiprazole augmentation in treatment-resistant late-life depression: an analysis of the IRL-GRey randomized clinical trial. JAMA Psychiatry 73, 329–336. https://doi.org/10.1001/jamapsychiatry.2015.3447.

Keefe, J.R., Webb, C.A., DeRubeis, R.J., 2016. In cognitive therapy for depression, early focus on maladaptive beliefs may be especially efficacious for patients with personality disorders. J. Consult. Clin. Psychol. 84, 353–364. https://doi.org/10.1037/ccp0000071.

Kemp, A.H., Gordon, E., Rush, A.J., Williams, L.M., 2008. Improving the prediction of treatment response in depression: integration of clinical, cognitive, psychophysiological, neuroimaging, and genetic measures. CNS Spectr. 13, 1066–1086. https://doi.org/10.1017/S1092852900017120.

Kertz, S.J., Koran, J., Stevens, K.T., Björgvinsson, T., 2015. Repetitive negative thinking predicts depression and anxiety symptom improvement during brief cognitive behavioral therapy. Behav. Res. Ther. 68, 54–63. https://doi.org/10.1016/j.brat.2015.03.006.

Kessler, R.C., 2018. The potential of predictive analytics to provide clinical decision support in depression treatment planning. Curr. Opin. Psychiatry 31, 32–39. https://doi.org/10.1097/YCO.0000000000000377.

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

Khazanov, G.K., Xu, C., Dunn, B.D., Cohen, Z.D., DeRubeis, R.J., Hollon, S.D., 2020. Distress and anhedonia as predictors of depression treatment outcome: a secondary analysis of a randomized clinical trial. Behav. Res. Ther. 125, 103507. https://doi.org/10.1016/j.brat.2019.103507.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Kudo, Y., Nakagawa, A., Wake, T., Ishikawa, N., Kurata, C., Nakahara, M., Nojima, T., Mimura, M., 2016. Temperament, personality, and treatment outcome in major depression: a 6-month preliminary prospective study. NDT 13, 17–24. https://doi.org/10.2147/NDT.S123788.

Laird, K.T., Lavretsky, H., St Cyr, N., Siddarth, P., 2018. Resilience predicts remission in antidepressant treatment of geriatric depression. Int. J. Geriatr. Psychiatry 33, 1596–1603. https://doi.org/10.1002/gps.4953.

Lee, J., Banerjee, J., Khobragade, P.Y., Angrisani, M., Dey, A.B., 2019. LASI-DAD study: a protocol for a prospective cohort study of late-life cognition and dementia in India. BMJ Open 9, e030300. https://doi.org/10.1136/bmjopen-2019-030300.

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

LoParo, D., Dunlop, B.W., Nemeroff, C.B., Mayberg, H.S., Craighead, W.E., 2025. Prediction of individual patient outcomes to psychotherapy vs medication for major depression. Npj Ment Health Res 4, 4. https://doi.org/10.1038/s44184-025-00119-9.

Lorenzo-Luaces, L., Rodriguez-Quintana, N., Bailey, A.J., 2020. Double trouble: do symptom severity and duration interact to predicting treatment outcomes in adolescent depression? Behav. Res. Ther. 131, 103637. https://doi.org/10.1016/j.brat.2020.103637.

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. Clin. Psychol. Sci. 7, 445–461. https://doi.org/10.1177/2167702618815466.

Maslej, M.M., Furukawa, T.A., Cipriani, A., Andrews, P.W., Sanches, M., Tomlinson, A., Volkmann, C., McCutcheon, R.A., Howes, O., Guo, X., Mulsant, B.H., 2021. Individual differences in response to antidepressants: a Meta-analysis of placebo-controlled randomized clinical trials. JAMA Psychiatry 78, 490–497. https://doi.org/10.1001/jamapsychiatry.2020.4564.

McGrath, J.J., Al-Hamzawi, A., Alonso, J., Altwaijri, Y., Andrade, L.H., Bromet, E.J., Bruffaerts, R., de Almeida, J.M.C., Chardoul, S., Chiu, W.T., Degenhardt, L., Demler, O.V., Ferry, F., Gureje, O., Haro, J.M., Karam, E.G., Karam, G., Khaled, S.M.,

Kovess-Masfety, V., Magno, M., Medina-Mora, M.E., Moskalewicz, J., Navarro-Mateu, F., Nishi, D., Plana-Ripoll, O., Posada-Villa, J., Rapsey, C., Sampson, N.A., Stagnaro, J.C., Stein, D.J., Ten Have, M., Torres, Y., Vladescu, C., Woodruff, P.W., Zarkov, Z., Kessler, R.C., WHO World Mental Health Survey Collaborators, 2023. Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. Lancet Psychiatry 10, 668–681. https://doi.org/10.1016/S2215-0366(23)00193-1.

Menchetti, M., Rucci, P., Bortolotti, B., Bombi, A., Scocco, P., Kraemer, H.C., Berardi, D., 2014. Moderators of remission with interpersonal counselling or drug treatment in primary care patients with depression: randomised controlled trial. Br. J. Psychiatry 204, 144–150. https://doi.org/10.1192/bjp.bp.112.122663.

Metts, A.V., Keilp, J.G., Kishon, R., Oquendo, M.A., Mann, J.J., Miller, J.M., 2018. Neurocognitive performance predicts treatment outcome with cognitive behavioral therapy for major depressive disorder. Psychiatry Res. 269, 376–385. https://doi.org/10.1016/j.psychres.2018.08.011.

Mihaljevic, S., Aukst-Margetic, B., Karnicnik, S., Vuksan-Cusa, B., Milosevic, M., 2016. Do spirituality and religiousness differ with regard to personality and recovery from depression? A follow-up study. Compr. Psychiatry 70, 17–24. https://doi.org/10.1016/j.comppsych.2016.06.003.

Mills, J.A., Suresh, V., Chang, L., Mayes, T., Croarkin, P.E., Trivedi, M.H., Strawn, J.R., 2022. Socioeconomic predictors of treatment outcomes among adults with major depressive disorder. PS 73, 965–969. https://doi.org/10.1176/appi.ps.202100559.

Moitra, M., Santomauro, D., Collins, P.Y., Vos, T., Whiteford, H., Saxena, S., Ferrari, A.J., 2022. The global gap in treatment coverage for major depressive disorder in 84 countries from 2000–2019: a systematic review and Bayesian meta-regression analysis. PLoS Med. 19, e1003901. https://doi.org/10.1371/journal.pmed.1003901.

Moradveisi, L., Huibers, M.J.H., Arntz, A., 2015. The influence of patients' attributions of the immediate effects of treatment of depression on long-term effectiveness of behavioural activation and antidepressant medication. Behav. Res. Ther. 69, 83–92. https://doi.org/10.1016/j.brat.2015.04.007.

Mulder, R.T., 2002. Personality pathology and treatment outcome in major depression: a review. AJP 159, 359–371. https://doi.org/10.1176/appi.ajp.159.3.359.

Newton-Howes, G., Tyrer, P., Johnson, T., Mulder, R., Kool, S., Dekker, J., Schoevers, R., 2014. Influence of personality on the outcome of treatment in depression: systematic review and Meta-analysis. J. Personal. Disord. 28, 577–593. https://doi.org/10.1521/pedi_2013_27_070.

Nguyen, K.P., Fatt, C.C., Treacher, A., Mellema, C., Cooper, C., Jha, M.K., Kurian, B., Fava, M., McGrath, P.J., Weissman, M., Phillips, M.L., Trivedi, M.H., Montillo, A.A., 2022. Patterns of pretreatment reward task brain activation predict individual antidepressant response: key results from the EMBARC randomized clinical trial. Biol. Psychiatry 91, 550–560. https://doi.org/10.1016/j.biopsych.2021.09.011.

OpenAI, 2023. New and Improved Embedding Model.

Patel, V., Chisholm, D., Rabe-Hesketh, S., Dias-Saxena, F., Andrew, G., Mann, A., 2003. Efficacy and cost-effectiveness of drug and psychological treatments for common mental disorders in general health care in Goa, India: a randomised, controlled trial. Lancet 361, 33–39. https://doi.org/10.1016/S0140-6736(03)12119-8.

Patel, V., Araya, R., Chowdhary, N., King, M., Kirkwood, B., Nayak, S., Simon, G., Weiss, H.A., 2008. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. Psychol. Med. 38, 221–228. https://doi.org/10.1017/S0033291707002334.

Patel, V., Chisholm, D., Parikh, R., Charlson, F.J., Degenhardt, L., Dua, T., Ferrari, A.J., Hyman, S., Laxminarayan, R., Levin, C., Lund, C., Medina Mora, M.E., Petersen, I., Scott, J., Shidhaye, R., Vijayakumar, L., Thornicroft, G., Whiteford, H., DCP MNS Author Group, 2016. Addressing the burden of mental, neurological, and substance use disorders: key messages from disease control priorities, 3rd edition. Lancet 387, 1672–1685. https://doi.org/10.1016/S0140-6736(15)00390-6.

Patel, V., Weobong, B., Weiss, H.A., Anand, A., Bhat, B., Katti, B., Dimidjian, S., Araya, R., Hollon, S.D., King, M., Vijayakumar, L., Park, A.-L., McDaid, D., Wilson, T., Velleman, R., Kirkwood, B.R., Fairburn, C.G., 2017. The healthy activity program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. Lancet 389, 176–185. https://doi.org/10.1016/S0140-6736(16)31589-6.

Perlis, R.H., Cusin, C., Fava, M., 2014. Proposed DSM-5 mixed features are associated with greater likelihood of remission in out-patients with major depressive disorder. Psychol. Med. 44, 1361–1367. https://doi.org/10.1017/S0033291712000281.

Perlman, K., Benrimoh, D., Israel, S., Rollins, C., Brown, E., Tunteng, J.F., You, R., You, E., Tanguay-Sela, M., Snook, E., Miresco, M., Berlim, M.T., 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.

Phillips, M.L., Chase, H.W., Sheline, Y.I., Etkin, A., Almeida, J.R.C., Deckersbach, T., Trivedi, M.H., 2015. Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches. Am. J. Psychiatry 172, 124–138. https://doi.org/10.1176/appi.ajp.2014.14010076.

Pozuelo, J.R., Lahiri, A., Singh, R.S.P., Kushwah, A., Khanduri, M., Shukla, A., Khan, A., G., S., Shende, V., Parashar, Y., Mehra, Y.K., Bhan, A., Kessler, R.C., Singla, D.R., Naslund, J.A., Choi, K.W., Cuijpers, P., DeRubeis, R., Herzallah, M.M., Lu, C., Smoller, J.W., VanderWeele, T.J., Rozatkar, A.R., Modak, T., Joel, M.M., Biswas, D., Atal, S., Kulsum, U., Hollon, S.D., Patel, V., 2025. Optimizing treatment for depression in primary care using psychotherapy versus antidepressant medication in a low-resource setting: protocol for the OptimizeD randomized controlled trial. BMC Psychiatry 25, 744. https://doi.org/10.1186/s12888-025-07030-9.

Puac-Polanco, V., Ziobrowski, H.N., Ross, E.L., Liu, H., Turner, B., Cui, R., Leung, L.B., Bossarte, R.M., Bryant, C., Joormann, J., Nierenberg, A.A., Oslin, D.W., Pigeon, W. R., Post, E.P., Zainal, N.H., Zaslavsky, A.M., Zubizarreta, J.R., Luedtke, A., Kennedy, C.J., Cipriani, A., Furukawa, T.A., Kessler, R.C., 2023. Development of a model to predict antidepressant treatment response for depression among veterans. Psychol. Med. 53, 5001–5011. https://doi.org/10.1017/S0033291722001982.

Quilty, L.C., Marshe, V., Lobo, D.S.S., Harkness, K.L., Müller, D.J., Bagby, R.M., 2017. Childhood abuse history in depression predicts better response to antidepressants with higher serotonin transporter affinity: a pilot investigation. Neuropsychobiology 74, 78–83. https://doi.org/10.1159/000453549.

Radkovsky, A., McArdle, J.J., Bockting, C.L.H., Berking, M., 2014. Successful emotion regulation skills application predicts subsequent reduction of symptom severity during treatment of major depressive disorder. J. Consult. Clin. Psychol. 82, 248–262. https://doi.org/10.1037/a0035828.

Rani, M., Bonu, S., Jha, P., Nguyen, S.N., Jamjoum, L., 2003. Tobacco use in India: prevalence and predictors of smoking and chewing in a national cross sectional household survey. Tob. Control. 12, e4. https://doi.org/10.1136/tc.12.4.e4.

Rathod, S., Pinninti, N., Irfan, M., Gorczynski, P., Rathod, P., Gega, L., Naeem, F., 2017. Mental health service provision in low- and middle-income countries. Health Serv. Insights 10, 1178632917694350. https://doi.org/10.1177/1178632917694350.

Rnic, K., Jung, Y.-E., Torres, I., Chakrabarty, T., LeMoult, J., Vaccarino, A.L., Morton, E., Bhat, V., Giacobbe, P., McInerney, S., Frey, B.N., Milev, R.V., Müller, D., Ravindran, A.V., Rotzinger, S., Kennedy, S.H., Lam, R.W., 2021. Association between discrepancy in objective and subjective cognitive abilities and treatment response in patients with major depressive disorder: a CAN-BIND-1 study report. J. Affect. Disord. 295, 1095–1101. https://doi.org/10.1016/j.jad.2021.09.002.

Rossom, R.C., Solberg, L.I., Vazquez-Benitez, G., Whitebird, R.R., Crain, A.L., Beck, A., Unützer, J., 2016. Predictors of poor response to depression treatment in primary care. PS 67, 1362–1367. https://doi.org/10.1176/appi.ps.201400285.

Saunders, R., Cape, J., Fearon, P., Pilling, S., 2016. Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. J. Affect. Disord. 197, 107–115. https://doi.org/10.1016/j.jad.2016.03.011.

Schilling, C., Weidner, K., Schellong, J., Joraschky, P., Pöhlmann, K., 2014. Patterns of childhood abuse and neglect as predictors of treatment outcome in inpatient psychotherapy: a typological approach. Psychopathology 48, 91–100. https://doi.org/10.1159/000368121.

Schilling, V.N.L.S., Zimmermann, D., Rubel, J.A., Boyle, K.S., Lutz, W., 2021. Why do patients go off track? Examining potential influencing factors for being at risk of psychotherapy treatment failure. Qual. Life Res. 30, 3287–3298. https://doi.org/10.1007/s11136-020-02664-6.

StataCorp, 2021. Stata: Release 17. Statistical Software.

Stikkelbroek, Y., Vink, G., Nauta, M.H., Bottelier, M.A., Vet, L.J.J., Lont, C.M., van Baar, A.L., Bodden, D.H.M., 2020. Effectiveness and moderators of individual cognitive behavioral therapy versus treatment as usual in clinically depressed adolescents: a randomized controlled trial. Sci. Rep. 10, 14815. https://doi.org/10.1038/s41598-020-71160-1.

Stiles-Shields, C., Kwasny, M.J., Cai, X., Mohr, D.C., 2014. Comorbid anxiety as a differential treatment predictor for TEL ephone versus face-to-face administered cognitive behavioral therapy for depression. Depress. Anxiety 31, 934–940. https://doi.org/10.1002/da.22323.

Stiles-Shields, C., Corden, M.E., Kwasny, M.J., Schueller, S.M., Mohr, D.C., 2015. Predictors of outcome for telephone and face-to-face administered cognitive behavioral therapy for depression. Psychol. Med. 45, 3205–3215. https://doi.org/10.1017/S0033291715001208.

Sumiyoshi, T., Watanabe, K., Noto, S., Sakamoto, S., Moriguchi, Y., Hammer-Helmich, L., Fernandez, J., 2021. Relationship of subjective cognitive impairment with psychosocial function and relapse of depressive symptoms in patients with major depressive disorder: analysis of longitudinal data from PERFORM-J. Neuropsychiatr. Dis. Treat. 17, 945–955. https://doi.org/10.2147/NDT.S288108.

Tait, J., Edmeade, L., Delgadillo, J., 2022. Are depressed patients' coping strategies associated with psychotherapy treatment outcomes? Psychol. Psychother. 95, 98–112. https://doi.org/10.1111/papt.12368.

Thiruchselvam, T., Dozois, D.J.A., Bagby, R.M., Lobo, D.S.S., Ravindran, L.N., Quilty, L. C., 2019. The role of outcome expectancy in therapeutic change across psychotherapy versus pharmacotherapy for depression. J. Affect. Disord. 251, 121–129. https://doi.org/10.1016/j.jad.2019.01.046.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267–288.

VanderWeele, T.J., 2022. Constructed measures and causal inference: towards a new model of measurement for psychosocial constructs. Epidemiology 33, 141–151. https://doi.org/10.1097/EDE.0000000000001434.

Veritas Health Innovation, 2024. Covidence Systematic Review Software.

Vinckier, F., Gourion, D., Mouchabac, S., 2017. Anhedonia predicts poor psychosocial functioning: results from a large cohort of patients treated for major depressive disorder by general practitioners. Eur. Psychiatry 44, 1–8. https://doi.org/10.1016/j.eurpsy.2017.02.485.

Wagner, S., Tadić, A., Roll, S.C., Engel, A., Dreimüller, N., Engelmann, J., Lieb, K., 2017. A combined marker of early non-improvement and the occurrence of melancholic features improve the treatment prediction in patients with major depressive disorders. J. Affect. Disord. 221, 184–191. https://doi.org/10.1016/j.jad.2017.06.042.

Wallert, J., Boberg, J., Kaldo, V., Mataix-Cols, D., Flygare, O., Crowley, J.J., Halvorsen, M., Ben Abdesslem, F., Boman, M., Andersson, E., Hentati Isacsson, N., Ivanova, E., Rück, C., 2022. Predicting remission after internet-delivered psychotherapy in patients with depression using machine learning and multi-modal data. Transl. Psychiatry 12, 357. https://doi.org/10.1038/s41398-022-02133-3.

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. J. Consult. Clin. Psychol. 88, 25–38. https://doi.org/10.1037/ccp0000451.
Weintraub, S., Dikmen, S.S., Heaton, R.K., Tulsky, D.S., Zelazo, P.D., Bauer, P.J., Carlozzi, N.E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N.A., Beaumont, J.L., Mungas, D., Nowinski, C.J., Richler, J., Deocampo, J.A., Anderson, J.E., Manly, J.J., Borosh, B., Havlik, R., Conway, K., Edwards, E., Freund, L., King, J.W., Moy, C., Witt, E., Gershon, R.C., 2013. Cognition assessment using the NIH toolbox. Neurology 80, S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded.
Whiston, A., Bockting, C.L.H., Semkovska, M., 2019. Towards personalising treatment: a systematic review and meta-analysis of face-to-face efficacy moderators of cognitive-behavioral therapy and interpersonal psychotherapy for major depressive disorder. Psychol. Med. 49, 2657–2668. https://doi.org/10.1017/S0033291719002812.
WHO, 2010. mhGAPintervention Guide for Mental, Neurological and Substance Use Disorders in Non- specialized Health Settings: Mental Health Gap Action Programme (mhGAP). WHO, Geneva.
Williams, L.M., Debattista, C., Duchemin, A.M., Schatzberg, A.F., Nemeroff, C.B., 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. https://doi.org/10.1038/tp.2016.61.
World Population, 2024. Madhya Pradesh Population 2025 [WWW Document]. URL. https://worldpopulationreview.com/regions/madhya-pradesh. accessed 5.11.25.
Worley, M.J., Tate, S.R., Granholm, E., Brown, S.A., 2014. Mediated and moderated effects of neurocognitive impairment on outcomes of treatment for substance dependence and major depression. J. Consult. Clin. Psychol. 82, 418–428. https://doi.org/10.1037/a0036033.
Yun, L., Fagan, M., Subramaniapillai, M., Lee, Y., Park, C., Mansur, R.B., McIntyre, R.S., Faulkner, G.E.J., 2020. Are early increases in physical activity a behavioral marker for successful antidepressant treatment? J. Affect. Disord. 260, 287–291. https://doi.org/10.1016/j.jad.2019.09.002.
Zilcha-Mano, S., Brown, P.J., Roose, S.P., Cappetta, K., Rutherford, B.R., 2019. Optimizing patient expectancy in the pharmacologic treatment of major depressive disorder. Psychol. Med. 49, 2414–2420. https://doi.org/10.1017/S0033291718003343.
Ziobrowski, H.N., Cui, R., Ross, E.L., Liu, H., Puac-Polanco, V., Turner, B., Leung, L.B., Bossarte, R.M., Bryant, C., Pigeon, W.R., Oslin, D.W., Post, E.P., Zaslavsky, A.M., Zubizarreta, J.R., Nierenberg, A.A., Luedtke, A., Kennedy, C.J., Kessler, R.C., 2022. Development of a model to predict psychotherapy response for depression among veterans. Psychol. Med. 1–10. https://doi.org/10.1017/S0033291722000228.
Zisook, S., Johnson, G.R., Tal, I., Hicks, P., Chen, P., Davis, L., Thase, M., Zhao, Y., Vertrees, J., Mohamed, S., 2019. General predictors and moderators of depression remission: a VAST-D report. Am. J. Psychiatry 176, 348–357. https://doi.org/10.1176/appi.ajp.2018.18091079.