

PROTOCOL

A protocol for the development of a core outcome set for adults with depression

C. Veal^{a,b,*}, K.R. Krause^a, E.I. Fried^c, A. Cipriani^{d,e,f}, P. Cuijpers^{g,h}, J. Downs^{i,j},
T.A. Furukawa^k, G. Gartlehner^{l,m}, S.D. Hollonⁿ, H. Levy-Soussan^{m,o}, G. Sahlem^p,
A. Tomlinson^d, S. Touboul^{j,q}, P. Ravaud^{a,b}, V.-T. Tran^{a,b}, A. Chevance^{a,b}

^aUniversité Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Center for Research in Epidemiology and Statistics (CRESS), F-75004, Paris, France

^bCentre d'Epidémiologie Clinique, AP-HP, Hôpital Hôtel Dieu, F-75004, Paris, France

^cDepartment Psychology, Clinical Psychology Unit, Leiden University, Leiden, The Netherlands

^dDepartment of Psychiatry, University of Oxford, Oxford, UK

^eOxford Precision Psychiatry Lab, NIHR Oxford Health Biomedical Research Centre, Oxford, UK

^fOxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK

^gDepartment of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

^hBabeş-Bolyai University, International Institute for Psychotherapy, Cluj-Napoca, Romania

ⁱPatient Representative, Faculty of Eating Disorders, Royal College of Psychiatrists, London, United Kingdom

^jCo-Researcher with Lived Experiences of Depression

^kDepartment of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, 606-8501, Japan

^lDepartment for Evidence-Based Medicine and Evaluation, University for Continuing Education Krems, Krems, Austria

^mCenter for Advanced Public Health Methods, RTI International, Research Triangle Park, NC, USA

ⁿDepartment of Psychology, Vanderbilt University, Nashville, USA

^oLa Maison Perchée, Paris, France

^pDepartment of Psychiatry and Behavioral Sciences, Duke University, Durham, North Carolina, USA

^qIndependent researcher, Paris, France

Accepted 20 December 2025; Published online 29 December 2025

Abstract

Background and Objective: Heterogeneous outcome measurement limits the comparison and combination of results from randomized controlled trials and observational studies aimed at evaluating therapeutic interventions for depression. We report here the protocol for the development of a Core Outcome Set (COS) for adults with depression.

Methods: Development will follow a multistep approach with: (1) generating outcome domains that matter to people with lived experiences of depression, health care professionals, and carers through a large online international survey using open-ended questions; (2) selecting domains based on the preferences of key interest holders through an international online preference elicitation survey; and (3) identifying relevant outcome measures with measurement properties considered sufficient through several systematic reviews conducted according to COnsensus-based Standards for the selection of health Measurement INstruments standards.

Discussion: The protocol describes a proof-of-concept approach to include large numbers of individuals from all key interest holder groups in COS development, which could be replicated in other conditions and contexts. © 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Core outcome set; Outcomes; Depression; Preference elicitation; Lived experiences

1. Introduction

1.1. Background

Depression affects approximately 280 million people around the world and is the second cause of years lived with disability [1]. A large number of randomized controlled

Funding: This study has received funding from Fondation de France.

* Corresponding author. Hôtel Dieu, 1, Place du Parvis de Notre-Dame, 75004, Paris, France.

E-mail address: christopher-james.veal@inserm.fr (C. Veal).

What is new?

Key findings

- Preference elicitation and aggregation methods can effectively capture and preserve the individual preferences of key interest holders, generating rich preference data that enhances the robustness and credibility of a consensus.

What this adds to what was known?

- Among a range of methods for eliciting interest holders' views about important outcomes, Core Outcome Set (COS) developers need to select those best aligned with their study's aims.
- Preference elicitation methods highlight which outcome domains matter most and to whom, enabling the fair capture of individual preferences from key interest holders.
- Granular preference data support the selection of outcomes relevant to subgroups of interest holders, broadening the COS's scope and applicability.

What is the implication and what should change now?

- COS developers must choose methods that best inform decision-makers and bolster the credibility of a consensus. Preference elicitation methods should be considered to ensure that consensus is built with methodological rigor and a diverse range of perspectives are fairly captured and preserved.

trials (RCTs) have been conducted to evaluate the efficacy of psychological, pharmacological, and other treatments for depression. However, the outcomes used in these trials may limit the interpretation of efficacy results for two reasons [2].

First, the outcome domains (ie, the distinct aspects and experiences of a condition) currently used in RCTs of depression do not fully represent what matters to people with lived experiences. In a previous study of 3003 people with depression, health care professionals, and carers, we identified 80 unique outcome domains representing what matters [3]. However, less than a quarter of these outcome domains were measured in late-stage RCTs of depression, while some (eg, mental pain, emotion regulation) were never measured [4]. Second, the outcome measures (ie, the instrument, scale, or method used to measure a specific outcome domain) are heterogeneous across RCTs of depression, with at most 40% of recent and ongoing RCTs using a common measure [4–6]. The heterogeneous measurement of outcomes that may only partially address what

matters to people with lived experiences limits the reliability of evidence on the efficacy of interventions.

To address these challenges, one solution is to develop a Core Outcome Sets (COS) for depression [6]. A COS is an agreed-on standardized set of outcomes that should be measured, as a minimum, in all trials for a specific clinical area and identifies what should be measured (ie, outcome domains) and how it should be measured (ie, outcome measures) [7]. Some mental health research funders, such as the National Institute of Mental Health (NIMH) and Wellcome Trust, endorse expert recommendations for a set of standardized measures that were developed under the auspices of the International Consortium for Health Outcomes Measurement (ICHOM), including the Patient Health Questionnaire-9 for depressive symptoms, the Generalized Anxiety Disorder-7 for anxiety symptoms, and the World Health Organization Disability Assessment Schedule 2.0 for functioning impairments [8,9]. However, as ICHOM sets are geared for use in clinical practice, this set of measures may not be best suited for clinical research [10]. Instead, trials evaluating treatment efficacy may favor other measures which capture more of what matters to people with lived experiences.

We therefore propose to develop a COS for depression for use in clinical research that incorporates measures that reflect what matters to people with depression, alongside the differing priorities of other key interest-holder groups, such as health care professionals, carers, and research funders [11,12]. This core set of outcomes promotes comparability across depression research, without precluding the inclusion of additional outcomes that address study-specific aims or the broader perspectives of other interest-holder groups.

2. Methods

This project is registered with the Core Outcome Measures in Effectiveness Trials (COMET) Initiative, and ongoing updates and findings will be published in the COMET database (<https://www.comet-initiative.org/Studies/Details/1105>). The reporting of this protocol follows the COS STANDARDIZED Protocol Items statement and the Guidance for Reporting Involvement of Patients and the Public 2 reporting checklists to improve reporting of patient and public involvement (PPI) in research (Supplementary materials 1 and 2) [13,14]. Fourteen key interest-holders, three of whom are people with lived experiences, participate in the study's steering committee, which will oversee the design and development of the COS and are also coauthors of this protocol.

Hereafter, in line with the terminology used in the scientific literature, the terms “people with depression” will be used to refer to those who are currently experiencing or have previously experienced depression. We acknowledge that lived experiences of depression include those who have

sought formal help from health care professionals as well as those who self-identify their experiences. We also acknowledge that cultural and linguistic differences shape how depression is described and labeled. By using the term “people with depression,” we do not intend to exclude those whose relevant experiences might be described or labeled differently.

2.1. Strategic vision

Three strategic methodological choices shape the development of our COS for depression. First, we envision the COS as applicable to diverse populations of people with depression and for a broad variety of interventions. Thus, we will prioritize the recruitment of large, international, and diverse samples of key interest-holders (ie, people with depression, health care professionals, and carers) to enhance the COS’ generalizability and credibility [15,16].

Second, we recognize the potential benefits of preference elicitation methods to better represent the preferences key interest-holders hold toward outcome domains. We will therefore use methods that fairly capture individual preferences at a large scale and international level and aggregate these into collective preferences according to subgroups of interest. This will limit token participation and improve representativeness of minorities [16].

Third, we anticipate the future need for comparable efficacy results from diverse data sources and methodologies. We will emphasize measurements that can be captured as part of routine outcome measurement or retrieved in routine care data (eg, electronic health record data) to enable the COS to be used in nonrandomized studies using these types of data sources.

2.2. Scope of the Core Outcome Set

We propose that the COS will cover adult populations (> 18 years old) with depression without the exclusion of any comorbidity. The COS aims to represent the diverse experiences of depression, which can vary depending on the social, cultural, and economic contexts of people and present as an isolated experience, recurrently (ie, depressive disorder or bipolar disorder), chronically (ie, persistent depressive disorder), and often associated with psychiatric and/or physical comorbidities [17–19]. The COS will not be limited by the treatment approach (eg, psychotherapies, drugs, neurostimulation, physical activity, complementary and alternative medicines) and will be usable in a broad scope of study designs that evaluate treatment efficacy (eg, RCTs, nonrandomized studies). The COS aims to create a shared framework across diverse studies of depression, enabling comparability and synthesis on common outcomes, without limiting the inclusion of further outcomes tailored to specific research interests or other interest-holder priorities.

2.3. Steps of the Core Outcome Set development

We outline three steps to develop a COS for adults with depression (Fig 1). Step 1 focuses on generating outcome domains that matter to people with lived experiences of depression. Step 2 focuses on eliciting the preferences of key interest-holders toward the outcome domains that most matter. Step 3 focuses on identifying outcome measures with satisfactory measurement properties that correspond with the outcome domains that most matter.

The steering committee will continue to review each step outlined in the protocol as the COS development project advances. All modifications to the original protocol will be considered in consultation with relevant experts and applied with the intention of using the best available methods at the time each step is completed.

2.3.1. Step 1: Generation of outcome domains

The first step of a COS is the generation of outcome domains that are relevant to key interest-holders [7]. We completed this first step with the PROCEED study (Participative Research on Outcomes and Core Expectations Elicitation for Depression) published in 2020 [3].

2.3.1.1. Design. We conducted a large online international survey with open-ended questions about the expected benefit of treatment for people with depression to identify which domains matter.

2.3.1.2. Population. In total, we included 1912 people with depression, 627 health care professionals, and 464 informal caregivers. All participants were adults (ie, > 18 years old) from a total of 52 countries.

2.3.1.3. Data collection. Participants answered open-ended questions on their perspectives of the expected benefits of depression treatment, developed during a qualitative pilot study involving people with depression, health care professionals, and carers. For example, people with depression were asked (among other questions): “For you, what is the most difficult aspect of depression to live with or endure?” Health care professionals were asked: “According to you, which criteria would be essential to measure in treatment efficacy studies for depression?” Carers were asked: “What do you consider most important to address in a depressed person?”

2.3.1.4. Data analysis. In total, 8183 open-ended answers were analyzed. We identified possible outcome domains with an inductive multiple-round qualitative content analysis. We used a mathematical model developed to assess the point of data saturation in surveys with open-ended questions, which showed that at least 90% of all possible domains had been identified for all participant groups [20]. In total, this step identified 80 candidate outcome domains that we considered suitable for further COS

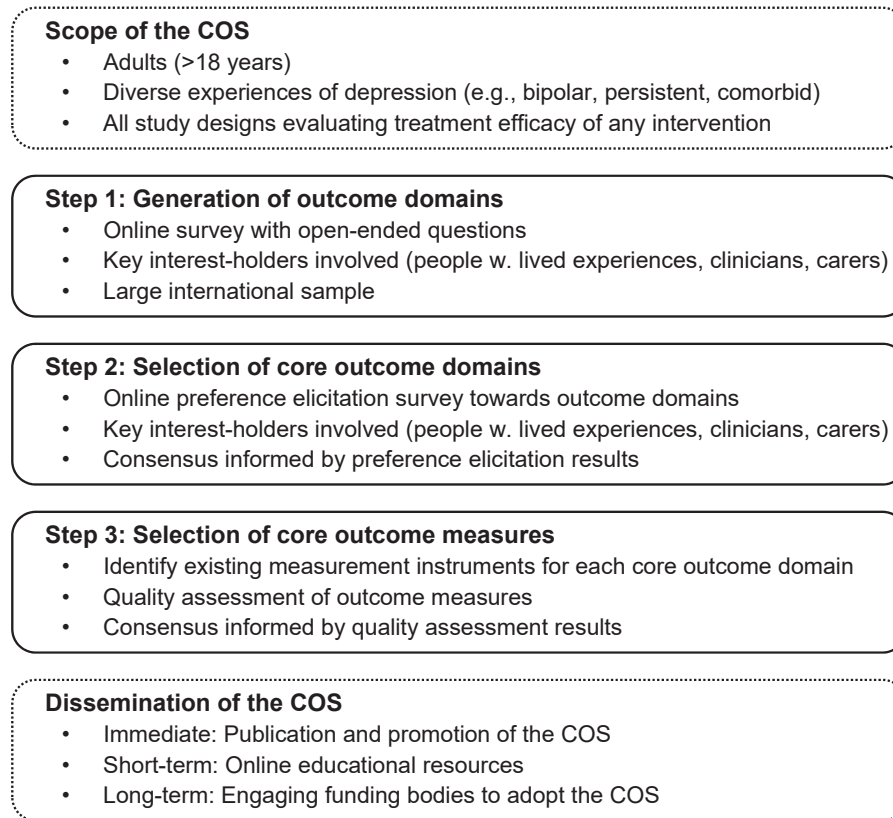


Figure 1. Proposed methods for the development of a COS for adults with depression. Adapted from the COMET Handbook 1.0 [7], Chevance et al [16], and Akinremi et al [34]. COS, Core Outcome Set.

development: 64 were related to symptoms and 16 were related to functioning impacts. These domains were described, in the “Depression dictionary” by a definition, a vignette illustrating the experience of the domain for a person with depression, and verbatim examples of the open-ended answers from which the domain was identified for the three participant groups [4].

2.3.2. Step 2: Selection of the core outcome domains

The second step of a COS is the selection of core outcome domains that are most preferred by key interest-holders [7].

2.3.2.1. Design. We will conduct a large online international survey with a preference elicitation task to identify which of the 80 outcome domains identified previously most matter to people with depression, health care professionals, and carers. These will be discussed in a consensus meeting with key interest-holders to define the final set of core outcome domains. Figure 2 presents the process for the selection of the core outcome domains.

2.3.2.2. Population. We will include adults (ie, > 18 years old) with access to the internet and speaking one of the five languages for which the survey will be made available (eg, English, French, German, Italian, and Spanish) as eligible

to participate. In addition, all participants must self-identify with one of the three interest-holder groups.

- People with lived experiences of depression: Self-report as having a history or ongoing experience of depression.
- Health care professionals: Self-report as having professional experience with people with depression (eg, general practitioners, psychiatrists, psychologists, nurses).
- Carers: Self-report as having a history or ongoing experience of supporting a person with depression (eg, family member, partner, housemate).

We aim for a sample size of 2000 people with lived experiences of depression with a history of or ongoing experience of depression; 500 health care professionals directly involved with people with depression (eg, clinicians, psychiatrists, general practitioners, nurses); and 300 carers who support a person with depression (eg, family members or friends). The sample size was estimated following Orme’s rule-of-thumb for conjoint analyses, which suggests a minimum of 200 participants per group for comparative studies [21]. We increased these estimates based on anticipated response rates and our prioritization of preferences from people with lived experiences. These estimates are also grounded in our prior experience conducting large

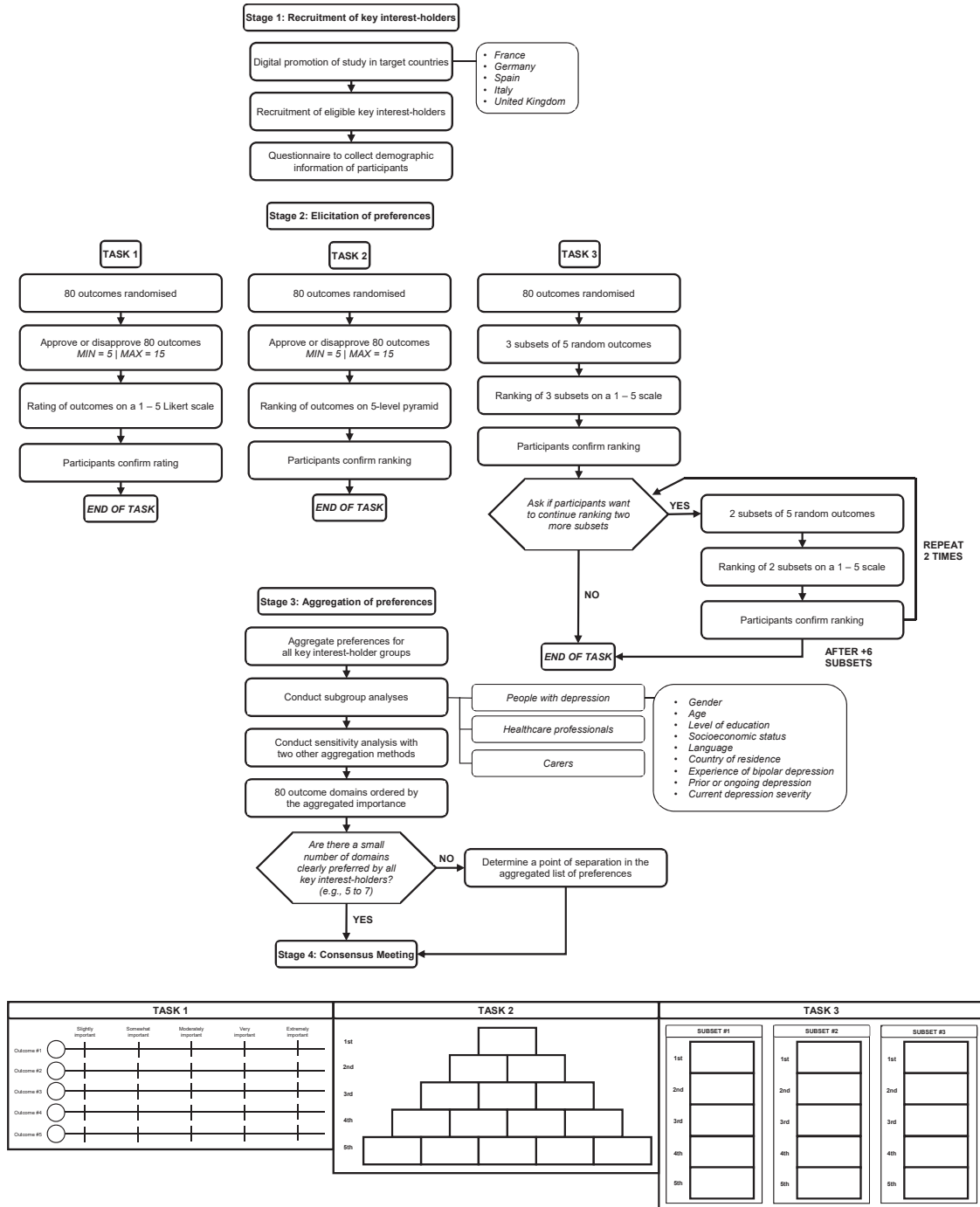


Figure 2. Flowchart for the selection of core outcome domains for the COS for adults with depression. COS, Core Outcome Set.

online international surveys with similar key interest-holder groups [3,22,23].

2.3.2.3. *Data collection.* The survey will be accessible on the online Platform for Research Online and CitizEn Science Surveys 24/7 and in five different languages to facilitate the inclusion of an international sample of participants.

The online survey will consist of two parts: a questionnaire to collect data on the characteristics of the participants and an interactive preference elicitation task to elicit participants’ preferences toward the 80 outcome domains. Regarding the questionnaire, we will collect data required to describe the population, in line with those collected in Step 1. Regarding the interactive preference elicitation task,

we will use one of three designs developed in collaboration with experts in preference elicitation and aggregation methods. A brief description of the three tasks is outlined below. The tasks were developed by leveraging the results of a review of methods to elicit preferences and an expert survey to explore their suitability for selecting core outcomes domains.

2.3.2.3.1. Task 1. First, participants follow an approval voting method to define an individual subset of preferred outcomes, by “approving” or “disapproving” the 80 outcomes. Outcomes will be presented randomly to avoid sequential order bias. We will impose the selection of a minimum of 5 outcomes and a maximum of 15 outcomes to facilitate data analysis. Then, participants quantify their preferences toward the outcomes included in their subset with a unipolar Likert scale with values ranging from 1 to 5 representing increasing levels of importance (eg, 1 = slightly important, 5 = extremely important). Each outcome will be presented vertically with corresponding draggable scales alongside them. Participants will validate their ratings before ending the task.

2.3.2.3.2. Task 2. First, participants define an individual subset of preferred outcomes, as described in the first step of task 1. For this task, we will require a selection of 15 outcomes in total to complete the ranking task. Then, participants quantify their preferences toward the outcomes included in their subset with a structured ranking pyramid of 5 levels of importance (eg, first place = most important, fifth place = least important). Each level has a corresponding number of positions on which outcomes can be placed (eg, first place = one position, fifth place = five positions). Participants can drag outcomes into a position on the pyramid and modify them until they best represent their preferred order. Participants will validate their rankings before ending the task.

2.3.2.3.3. Task 3. First, participants will rank three randomized subsets composed of five outcomes on a scale with 5 levels of importance (eg, first place = most important, fifth place = least important). Each level of importance will have only one corresponding position. For each subset, participants can drag outcomes into position on the scale and modify them until they best represent their preferred order. Participants will validate their rankings before proceeding. Then, participants will be asked if they want to rate an additional two subsets of outcomes (eg, +10 randomly selected outcomes). If participants decline, the preference elicitation task will end. If participants accept, they will rank an additional two randomized subsets using the same procedure. The proposal of ranking two additional subsets of outcomes will be proposed a total of three times before the preference elicitation task ends.

Before conducting the study, we will pilot test the three preference elicitation tasks with a small sample of French-speaking people with depression, health care professionals, and carers. Semistructured interviews with participants will capture feedback necessary to improve the

comprehensibility, feasibility, and acceptability of the preference elicitation task and select the most suitable method for the online survey. A dedicated protocol, outlining the chosen preference method in detail, will be registered on Open Science Framework before the inclusion of the first participant.

2.3.2.4. Data analysis. We will employ descriptive statistics to describe the sample of participants. We will engage experts in preference aggregation methods to choose the primary method to analyze the preferences of each interest-holder group (ie, people with depression, health-care professionals, carers). Sensitivity analyses will involve using different methods to assess how the choice of methods can affect results. We will analyze preferences in subgroups of people with depression defined by gender (eg, female, male, other), level of education, socioeconomic status, language, countries of residence, age (eg, young adults under 25 and elderly people over 65), and with respect to health characteristics (eg, self-reported diagnosis of bipolar disorder, ongoing episode vs history of depression, severe vs mild depression).

The output of this step will be a list of preferred outcome domains, ordered by the aggregated importance given to each of the 80 candidate outcome domains and combining all interest-holder groups together. Additional lists of preferred outcome domains will be obtained for each of the interest-holder groups and subgroups of people with depression. At this stage, we imagine two potential scenarios for the output of this study. In the first scenario, there is a small number of outcome domains clearly preferred by interest-holders and consistent across subgroups. Previous research from Outcome Measures in Rheumatology (OMERACT) indicates the ideal number of core outcome domains is between five and seven [24]. In the second scenario, there are more preferred outcome domains. In this instance, the outcome domains considered to be most preferred by interest-holder groups (ie, those above a clear separation point) will be discussed in a consensus meeting.

2.3.2.5. Consensus. We will conduct an online consensus meeting with a small number of key interest-holders to review and discuss the findings of the preference elicitation and aggregation study. The goal of the consensus meeting will be to decide which of the preferred outcome domains will be included in the core outcome domain set.

The consensus meeting will include between 15 and 20 representatives from key interest-holder groups and will use a nominal group technique guided by a trained facilitator. Participants will include members of the study’s steering committee, participants involved in the preference elicitation survey (eg, people with depression, health care professionals, carers), and other key interest-holders of depression research (eg, research funders and trialists). The consensus meeting will be held via videoconference to support international participation. A facilitator will present the lists

of preferred outcome domains and their corresponding descriptions, as defined in the Depression Dictionary [4]. The preferred outcome domains with the greatest consensus among all interest-holders will be selected as the core outcome domains. If some outcome domains appear to be clear priorities for specific subgroups (eg, those with bipolar disorder), we will chart these as additional outcomes, similar to the onions diagram of OMERACT [25].

2.3.3. Step 3: Selection of the core outcome measures

The third step of a COS is the selection of core outcome measures that best measure the core outcome domains [7].

2.3.3.1. Design. We will conduct a Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) review of outcome measures frequently used in RCTs of depression and match these measures to the

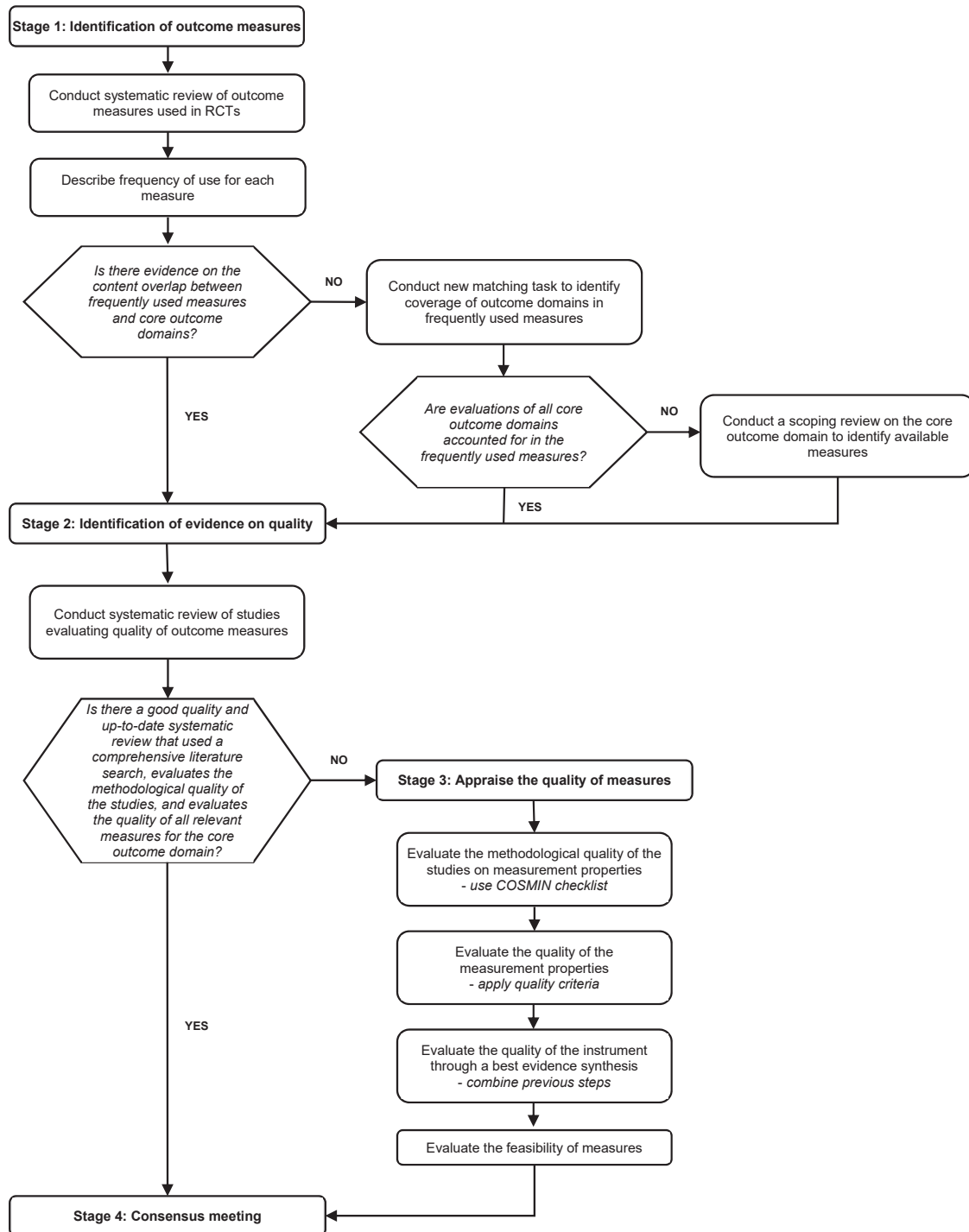


Figure 3. Flowchart for the selection of core outcome measures for the COS for adults with depression. Adapted from the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Guideline for selecting outcome measurement instruments for COS [26]. COS, Core Outcome Set; RCT, randomized controlled trial.

domains of PROCEED. We will identify frequently used measures in RCT registries, using the same methods outlined in our previous systematic review [4]. Measures with the highest quality will be discussed in a consensus meeting with key interest-holders to define the final set of core outcome measures. Figure 3 presents the process for the selection of the core outcome measures.

2.3.3.2. Data collection. We will search records of recent and ongoing RCTs to identify frequently used outcome measures in five registries: International Clinical Trials Registry Platform, [ClinicalTrials.gov](https://www.clinicaltrials.gov), EU Clinical Trials Register, and EU Clinical Trials Information System, and Metapsy. We consider frequently used measures to be those used by more than five trials.

Then, we will evaluate to what extent these measures cover the outcomes that matter, using the results of our previous matching task, which charted the overlap of the 80 outcome domains across 25 of the most-frequently used outcome measures in RCTs of depression [4]. In total, 91% of the outcomes were charted to at least one measure. If there is no evidence for a measure, we will conduct a new matching task using the same procedure. If the core outcome domains are not represented by any of the frequently used measures, we will conduct scoping reviews in two literature databases (eg, PubMed and Embase) to identify existing measures. For example, Charvet et al [26] who searched for all published standardized measures of mental pain and identified the quality of 10 self-reported outcome measures.

For each core outcome domain, we will obtain a list of candidate outcome measures described by their frequency of use in recent and ongoing RCTs and the coverage of the 80 candidate outcome domains included in the measure. For example, the Pittsburgh Sleep Quality Index would be described as a measure of “sleep” used in 5% of RCTs of depression that measures the outcome domains “hypersomnia,” “insomnia,” and “disturbed sleep,” in addition to five other candidate outcome domains (eg, interest, anhedonia, fatigue) [4].

Finally, we will search for evidence regarding the quality of measures for each core outcome domain, according to the COSMIN recommendations [27]. If the evidence on measures is outdated, of poor quality, or there is no evidence on the quality of measures, we will appraise the outcome measures.

2.3.3.3. Data analysis. We will establish a working group, including people with depression, to assess the quality of each outcome measure using the COSMIN guidelines [28]. We will evaluate the validity, reliability, responsiveness, and interpretability of each measure, as well as the methodological quality of the underlying psychometric evidence, using the COSMIN taxonomy and methodologies for the study [29–32].

Within the working group, each evaluator will independently rate the retrieved information about the development

(COSMIN box 1), content validity (COSMIN box 2), and, if relevant, other measurement properties (COSMIN boxes 3–10). The overall rating of the aspects is determined using a rating algorithm summarizing the quality of evidence as high, moderate, low, or very low. Disagreements among evaluators will be resolved by consultation with experts in the assessment of outcome measures. We will maintain the worst-score-counts method, which takes the lowest rating of any aspect for the overall rating, as good-quality aspects cannot compensate for poor quality [33].

For each core outcome domain, we will obtain a list of measures with evidence on its frequency of use in RCTs, coverage of the 80 candidate outcome domains, and quality of its measurement properties.

2.3.3.4. Consensus. We will conduct an online consensus meeting with a small number of key interest-holders to review and discuss the results of the identification, evidence on quality, and appraisal of the candidate outcome measures. We will use the same format of the consensus meeting outlined in Step 2. The goal of the consensus meeting is to decide which of the corresponding outcome measures will be included in the core outcome measurement set.

During the consensus meeting, the facilitator will present the findings, such as the frequency of use in RCTs, the reporting method of the outcome measure (eg, patient-reported, clinician-reported), which candidate outcome domains are included in the measure, and quality appraisal of the measures. This information will help participants involved in the consensus meeting to decide whether the core outcome measures will be restricted to one measure for each core outcome domain or whether a few items within a measure are sufficient for measuring the core outcome domain. We consider the core outcome measures to be those with the greatest consensus among all interest-holders. If some outcome measures are suitable for evaluating the additional outcomes defined for specific subgroups (eg, those with bipolar disorder), we will chart these outcome measures separately.

2.4. Dissemination

The COS will define the outcome domains key interest-holders consider to be most important for adults with depression and the corresponding outcome measures that can be used to best measure the outcome domains. Findings from each step of the COS development project will be published in peer-reviewed international journals. The final report of the COS for adults with depression will be registered with the COMET database of COS projects.

Following the recommendations for COS dissemination suggested by Akinremi et al [34], we will disseminate the COS by contacting key interest-holders directly to advocate for its use in RCTs, provide online COS resources (eg, publication of the Depression Dictionary, adapted preference elicitation of outcome domains for clinical practice),

promote the COS and its core domains on social media (eg, posts promoting the COS and relevant outcomes for people with depression), develop educational programs (eg, COS development in psychiatry and psychology), and collaborate with funding bodies to promote the inclusion of standardized outcomes for depression (eg, the NIMH and the Wellcome Trust via the Common Measures in Mental Health Initiative, the World Health Organization).

Finally, the research team will continuously work with PPI initiatives and contributors. We will create resources specifically targeted toward people with depression that can be freely accessed online. We will also continue to work with PPI groups to build knowledge on relevant outcomes of depression by developing presentations, conducting information sessions, and improving the resources available to these groups on the outcome domains and outcome measures used in depression studies.

3. Discussion

Currently, no COS for adults with depression exists for use in clinical research. Developing one can help efficacy studies measure more of what most matters to key people with lived experiences of depression. Consistently measuring these outcomes across a variety of study designs can facilitate the translation of efficacy results for treatments of depression.

A key strength of our proposed approach is the inclusion of a large international sample of interest-holders, including both people with depression and other key interest-holders (eg, health care professionals, carers, methodologists, editors) throughout the entire development of the COS. We extend beyond the inclusion of minimal number of people with lived experiences by demonstrating methods that can meaningfully involve thousands of from multiple countries, as in the PROCEED study for generating outcome domains. Large international samples enhance the granularity of information we can generate on outcome domains that can support the measurement of what most matters to different types of people with depression.

In addition, the study advances the methodology for selecting outcome domains by incorporating innovative preference elicitation and aggregation methods that are not commonly used in COS development studies. These methods keep preference elicitation and aggregation separate from the consensus process, allowing for the effective capture and preservation of key interest-holders' individual preferences, while also enhancing the credibility of the consensus process by generating robust and granular preference data. This granular information can bolster the usefulness of the COS by supporting researchers to select common outcomes across different designs, interventions, or types of experiences.

Finally, our approach can be extended to meet the needs of emerging designs in the field of comparative effectiveness research, which holds the advantage of facilitating the

translation of evidence in emerging designs efficacy studies, such as emulated target trials, which takes into consideration the outcomes measured by observational data [35–37]. We conceptualize the COS as a dynamic tool that requires periodic revision and extension for emerging study designs, which may occur due to potential changes in the definition of depressive disorders, novel therapeutics, or further methodological innovations.

Standardizing the measurement of the outcome domains that most matter to those with lived experiences of depression increases the potential to identify and compare the effectiveness of treatments for depression. By positioning the COS as a minimum set of outcomes to be measured, studies' compliance with pre-existing regulatory frameworks is supported, and the findings of previous depression research are preserved.

CRedit authorship contribution statement

C. Veal: Writing – review & editing, Writing – original draft, Project administration, Methodology, Conceptualization. **K.R. Krause:** Writing – review & editing, Conceptualization. **E.I. Fried:** Writing – review & editing, Conceptualization. **A. Cipriani:** Writing – review & editing, Conceptualization. **P. Cuijpers:** Writing – review & editing, Conceptualization. **J. Downs:** Writing – review & editing, Conceptualization. **T.A. Furukawa:** Writing – review & editing, Conceptualization. **G. Gartlehner:** Writing – review & editing, Conceptualization. **S.D. Holton:** Writing – review & editing, Conceptualization. **H. Levy-Soussan:** Writing – review & editing, Conceptualization. **G. Sahlem:** Writing – review & editing, Conceptualization. **A. Tomlinson:** Writing – review & editing, Conceptualization. **S. Touboul:** Writing – review & editing, Conceptualization. **P. Ravaud:** Writing – review & editing, Conceptualization. **V.-T. Tran:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **A. Chevance:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Declaration of competing interest

AT has received research, educational and consultancy fees from INCiPiT (Italian Network for Pediatric Trials), Angelini Pharma, Takeda and acted as a clinical advisor for Akrivia Health.

AC is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006), by the NIHR Oxford and Thames Valley Applied Research Collaboration, by the NIHR Oxford Health Biomedical Research Center (grant

NIHR203316) and by the Wellcome Trust (GALENOS Project). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health and Social Care. He has received research, educational and consultancy fees from INCiPiT (Italian Network for Pediatric Trials), CARIPLO Foundation, Lundbeck and Angelini Pharma.

GLS has collaborated with MagVenture and MECTA as part of investigator-initiated trials, consults for and has equity in the company Trial Catalyst, and has provided consultation to Indivior.

TAF reports personal fees from Boehringer-Ingelheim, DT Axis, Kyoto University Original, Shionogi, SONY and UpToDate, and a grant from Shionogi, outside the submitted work; In addition, TAF has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. There are no competing interests for any other author.

Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jclinepi.2025.112119>.

Data availability

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

References

- [1] Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019). Institute for Health Metrics and Evaluation (IHME). 2020. Available at: <https://vizhub.healthdata.org/gbd-results>. Accessed March 27, 2024.
- [2] Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database — update and key issues. *N Engl J Med* 2011;364(9):852–60. <https://doi.org/10.1056/NEJMsa1012065>.
- [3] Chevance A, Ravaud P, Tomlinson A, Le Berre C, Teufer B, Touboul S, et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. *Lancet Psychiatry* 2020;7:692–702.
- [4] Veal C, Tomlinson A, Cipriani A, Bulteau S, Henry C, Müh C, et al. Heterogeneity of outcome measures in depression trials and the relevance of the content of outcome measures to patients: a systematic review. *Lancet Psychiatry* 2024;11(4):285–94. [https://doi.org/10.1016/S2215-0366\(23\)00438-8](https://doi.org/10.1016/S2215-0366(23)00438-8).
- [5] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383(9912):156–65. [https://doi.org/10.1016/S0140-6736\(13\)62229-1](https://doi.org/10.1016/S0140-6736(13)62229-1).
- [6] Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy* 2012;17(1):1–2. <https://doi.org/10.1258/jhsrp.2011.011131>.
- [7] Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. *Trials* 2017;18(3):280. <https://doi.org/10.1186/s13063-017-1978-4>.
- [8] Farber GK, Gage S, Kemmer D, White R. Common measures in mental health: a joint initiative by funders and journals. *Lancet Psychiatry* 2023;10(6):465–70. [https://doi.org/10.1016/S2215-0366\(23\)00139-6](https://doi.org/10.1016/S2215-0366(23)00139-6).
- [9] Obbarius A, van Maasackers L, Baer L, Clark DM, Crocker AG, de Beurs E, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. *Qual Life Res* 2017;26(12):3211–25. <https://doi.org/10.1007/s11136-017-1659-5>.
- [10] Patalay P, Fried EI. Editorial perspective: prescribing measures: unintended negative consequences of mandating standardized mental health measurement. *J Child Psychol Psychiatry* 2021;62(8):1032–6. <https://doi.org/10.1111/jcpp.13333>.
- [11] Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22(8):1889–905. <https://doi.org/10.1007/s11136-012-0344-y>.
- [12] Schünemann HJ, Tunis S, Nieuwlaat R, Wiercioch W, Baldeh T. Controversy and debate series on core outcome sets: the SOLAR (Standardized Outcomes Linking Across StakeholderS) system and hub and spokes model for direct core outcome measures in health care and its relation to GRADE. *J Clin Epidemiol* 2020;125:216–21. <https://doi.org/10.1016/j.jclinepi.2020.05.007>.
- [13] Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core outcome Set-STandardised protocol items: the COS-STAP statement. *Trials* 2019;20(1):116. <https://doi.org/10.1186/s13063-019-3230-x>.
- [14] Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. *BMJ* 2017;358:j3453. <https://doi.org/10.1136/bmj.j3453>.
- [15] Frith L. Democratic justifications for patient public involvement and engagement in health research: an exploration of the theoretical debates and practical challenges. *J Med Philos Forum Bioeth Philos Med* 2023;48(4):400–12. <https://doi.org/10.1093/jmp/jhad024>.
- [16] Chevance A, Tran VT, Ravaud P. Improving the generalizability and the credibility of Core Outcome Sets (COS) by a large and international participation of diverse stakeholders. *J Clin Epidemiol* 2020;125:206–212.e1.
- [17] Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. *J Affect Disord* 2015;172:96–102. <https://doi.org/10.1016/j.jad.2014.10.010>.
- [18] Berk M, Köhler-Forsberg O, Turner M, Penninx BWJH, Wrobel A, Firth J, et al. Comorbidity between major depressive disorder and physical diseases: a comprehensive review of epidemiology, mechanisms and management. *World Psychiatry* 2023;22(3):366–87. <https://doi.org/10.1002/wps.21110>.
- [19] Sun S, Kuja-Halkola R, Faraone SV, D'Onofrio BM, Dalsgaard S, Chang Z, Larsson H. Association of psychiatric comorbidity with the risk of premature death among children and adults with attention-Deficit/Hyperactivity disorder. *JAMA Psychiatry* 2019;76(11):1141–9. <https://doi.org/10.1001/jamapsychiatry.2019.1944>.
- [20] Tran VT, Porcher R, Tran VC, Ravaud P. Predicting data saturation in qualitative surveys with mathematical models from ecological research. *J Clin Epidemiol* 2017;82:71–78.e2. <https://doi.org/10.1016/j.jclinepi.2016.10.001>.
- [21] Orme BKIn: *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 2. Madison WI, USA: Research Publishers LLC; 2010.
- [22] Chevance A, Tomlinson A, Ravaud P, Touboul S, Henshall C, Tran VT, Cipriani A. Important adverse events to be evaluated in antidepressant trials and meta-analyses in depression: a large international preference study including patients and healthcare

- professionals. *Evid Based Ment Health* 2022;25(e1):e41–8. <https://doi.org/10.1136/ebmental-2021-300418>.
- [23] Malmartel A, Ravaud P, Tran VT. A methodological framework allows the identification of personomic markers to consider when designing personalized interventions. *J Clin Epidemiol* 2023;159:235–45. <https://doi.org/10.1016/j.jclinepi.2023.06.003>.
- [24] Beaton D, Maxwell L, Grosskleg S, Shea B, Tugwell P. The OMERACT handbook Version 2.1 [updated April 2021]. 2021. Available at: <https://omeract.org/handbook/>. Accessed October 22, 2024.
- [25] Mackie SL, Twohig H, Neill LM, Harrison E, Shea B, Black RJ, et al. The OMERACT core domain set for outcome measures for clinical trials in Polymyalgia Rheumatica. *J Rheumatol* 2017;44(10):1515–21. <https://doi.org/10.3899/jrheum.161109>.
- [26] Charvet C, Boutron I, Morvan Y, Le Berre C, Touboul S, Gaillard R, et al. How to measure mental pain: a systematic review assessing measures of mental pain. *Evid Based Ment Health* 2022;25(4):e4. <https://doi.org/10.1136/ebmental-2021-300350>.
- [27] Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set” — a practical guideline. *Trials* 2016;17(1):449. <https://doi.org/10.1186/s13063-016-1555-2>.
- [28] Consensus-based Standards for the selection of health Measurement INstruments Initiative (COSMIN). I want to select the most suitable outcome measurement instrument • COSMIN. COSMIN. <https://www.cosmin.nl/finding-right-tool/select-best-measurement-instrument/>. Accessed May 3, 2024.
- [29] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27(5):1147–57. <https://doi.org/10.1007/s11136-018-1798-3>.
- [30] Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf. Accessed June 26, 2024.
- [31] Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27(5):1159–70. <https://doi.org/10.1007/s11136-018-1829-0>.
- [32] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27(5):1171–9. <https://doi.org/10.1007/s11136-017-1765-4>.
- [33] Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21(4):651–7. <https://doi.org/10.1007/s11136-011-9960-1>.
- [34] Akinremi A, Turnbull AE, Chessare CM, Bingham CO, Needham DM, Dinglas VD. Delphi panelists for a core outcome set project suggested both new and existing dissemination strategies that were feasibly implemented by a research infrastructure project. *J Clin Epidemiol* 2019;114:104–7. <https://doi.org/10.1016/j.jclinepi.2019.05.026>.
- [35] Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183(8):758–64. <https://doi.org/10.1093/aje/kwv254>.
- [36] Szmulewicz AG, Wanis KN, Perlis RH, Hernández-Díaz S, Öngür D, Hernán MA. Emulating a target trial of dynamic treatment strategies for major depressive disorder using data from the STAR*D randomized trial. *Biol Psychiatry* 2023;93(12):1127–36. <https://doi.org/10.1016/j.biopsych.2022.09.028>.
- [37] Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA* 2022;328(24):2446–7. <https://doi.org/10.1001/jama.2022.21383>.