

Measurement properties of the Revised Child Anxiety and Depression Scale 25 items (RCADS-25) child and parent versions across purposes and populations worldwide: Protocol for a Living Systematic Review

Yasmeen Aldouri, Sorina Andrei, Corentin J Gosling, Inika Khosla, Cinthya Ayerim Lucio Garcia, Clara Miguel, Lidwine Mokkink, Alan Ricardo Patlan-Hernandez, Jose T Ramos-Rojas, Kevin John Samuel, Colin Sidre, Abiramy Srirangan, Fatahiya Sumalia, Pierre-Felix Tellier, Gull Zareen, Karolin Rose Krause, Astrid Chevance, Suneeta Monga

GENERAL INFORMATION:

Type and method of the review: Systematic review and meta-analysis conducted following the COSMIN guideline for systematic reviews of patient-reported outcome measures (PROMs), and the PRISMA-COSMIN guideline for reporting systematic reviews of outcome measurement instruments (OMIs).

Condition or domain being studied: This systematic review focuses on anxiety and depression.

Language: English.

Country: France.

Funding sources/sponsors: This work is part of the COLIVE-MH project funded by Wellcome under contract number C-013403.

Conflicts of interest: All authors declare that they have no conflict of interest.

Ethical approval and consent to participate: This study synthesizes and analyzes data from already published studies, therefore ethical approval is not required.

Other registration details: None.

Dissemination plans: We plan to publish the review findings in a peer-reviewed journal.

Details of any existing review of the same topic by the same authors: None.

Current review status: Ongoing.

Additional information: None.

Protocol updated on 10th February 2026

Acknowledgement: We thank Jessie Cunningham, BA (Hons); MSt (The Hospital for Sick Children) for peer review of the submitted search strategy, and Catherine Auzimour for the project management.

Introduction

Anxiety and depressive symptoms in children and adolescents are common and frequently co-occur together. Development of these symptoms during childhood or adolescents is associated with impaired functioning across multiple areas including academic, social and family functioning. Additionally, the presence of these symptoms often results in elevated service use, and increased risk of ongoing, persistent mental health problems. Brief self-report instruments are therefore widely used to assess symptoms of anxiety or depression and to monitor change over time in children and youth. For these instruments to inform clinical and research decisions, their measurement properties must be well established across developmental stages, informants, and settings.

The Revised Child Anxiety and Depression Scale (RCADS-47) is a Diagnostic and Statistical Manual (DSM) oriented questionnaire designed to assess anxiety and depressive symptoms in children and adolescents (Chorpita et al., 2000). The RCADS-47 has both a child-report and parent-report version. To improve feasibility in routine care and large-scale applications, a 25-item short form (RCADS-25) of the child version was developed (Ebesutani et al., 2012). This shortened version of the RCADS consists of a 15-item Anxiety Total scale measuring a single “broad anxiety” dimension and a 10-item Depression Total scale derived from the original RCADS-47. Subsequent work has extended this short child-report form to include a parent-report version as well (Ebesutani et al., 2017). The RCADS-25 is now used in a variety of contexts, including as a brief screening tool and as an outcome measure in clinical and epidemiological studies.

Despite the broad use of various patient-reported outcome measures (PROMs) in mental health, the overall evidence on the reliability, validity and other important measurement properties across PROMs remains fragmented (Bear et al., 2020; Mew et al., 2020). For example, a recent COSMIN review of 12 PROMs that measure depression symptom severity in adolescents with depression did not find that any of the 12 PROMs were appropriate for use (Monga et al., 2025). This variability ultimately hinders clinicians, researchers, and other stakeholders to identify the most robust and up-to-date evidence to support informed use of the PROM.

Furthermore, inconsistencies across results of primary studies and reviews may reflect heterogeneity in how measurement properties are defined, operationalised, and assessed, as well as an absence of shared terminology and methodological standards in health measurement (Mokkink et al., 2009, 2010). Such variation further contributes to divergent findings and limits comparability across studies, highlighting the need for synthesis based on consistent and transparent methodological standards.

To address these challenges, the present study adopts a living systematic review approach to examine the measurement properties of the RCADS-25 child and parent versions. By continuously incorporating newly published evidence using a standardised methodological

framework, the review aims to maintain an up-to-date and coherent evidence-base as the literature evolves. The review will be conducted using the taxonomy (Mokkink et al., 2010) and guidance developed by the COSMIN initiative (COnsensus-based Standards for the selection of health Measurement INstruments) (Mokkink et al., 2024), which provides standards for evaluating study and instrument quality. In parallel, the results will be integrated into an open-access, continuously updated platform that presents the evidence in a structured and accessible format to support transparency, comparison across contexts, and knowledge translation.

The findings from this review will inform users about the measurement properties of the RCADS-25 (child and parent versions) across purposes of use, populations, and settings. Study results will also help determine whether reported inconsistencies reflect systematic and interpretable patterns across groups and contexts. By applying consistent methodological standards and updating the evidence base on an ongoing basis, this review will improve the comparability, transparency, and cumulative value of evidence on the measurement properties of the RCADS-25 (child and parent versions).

Methods

1. Objectives

1. The primary objective of this project is twofold:
 - (i) to systematically identify, appraise, and synthesize all available evidence on the measurement properties, as defined by the COSMIN framework, of the RCADS-25 (child and parent versions) in measuring depressive and separately anxiety symptoms across purposes (e.g., outcome measure of treatment effect, screening) and populations worldwide, using a living evidence synthesis approach; and
 - (ii) to disseminate this evidence through an open access, regularly updated and user-friendly web-based platform to support researchers, clinicians, people with lived experiences and other interest-holders in the informed selection and use of the scale.
2. The secondary objectives are:
 - (i) to evaluate the feasibility of implementing the RCADS-25 child and parent version in clinical and research settings;
 - (ii) to describe the interpretability of its scores; and
 - (iii) to identify and highlight key research gaps.

This protocol is part of COLIVE-MH, a broader project designed to build a regularly updated platform on the measurement properties of the four patient-reported outcome measures (PROMs) recommended by the Common Measures in Mental Health Science Initiative (CMMHS) (Farber et al., 2023). These include Revised Child Anxiety and Depression Scale 25-item (RCADS-25); the Patient Health Questionnaire 9-item scale (PHQ-9), the Generalized Anxiety Disorder 7-item scale (GAD-7), and the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). It follows the methodological framework described in the COLIVE-MH Master

Protocol. The project involves international teams of researchers in clinical epidemiology, psychiatry and anthropology, and co-researchers with lived experience.

2. Study design and reporting guidelines

The protocol of this systematic review follows the PRISMA-P (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols) guidelines (Moher et al., 2015). It will be prospectively registered on both Zenodo (<https://zenodo.org/>) and on PROSPERO (<https://www.crd.york.ac.uk/prospero/>). We will list protocol deviations and amendments in the Supplementary materials of any future publications. The publication of the systematic review results will be reported following the PRISMA-COSMIN for OMIs reporting guidelines (Elsman et al., 2024) and the PRISMA-LSR extension given the living nature of the review (Akl et al., 2024). We will report the involvement of experts with lived experience following the GRIPP-2 framework (Staniszewska et al., 2017).

The development and peer review of the search strategy will follow the PRESS (Peer Review of Electronic Search Strategies) 2015 guidelines (McGowan et al., 2016). The systematic reviews of measurement properties will be conducted according to the COSMIN 2.0 guidelines (Mokkink et al., 2024), which includes recommendations for measurement properties evidence synthesis and quality appraisal. In addition, the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Deeks et al., 2023) and the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) guidelines (Schueller et al., 2012) will be consulted for additional guidance when evaluating the RCADS-25 performance as a screening or diagnostic tool.

3. Formulation of the research question

Construct: The present review will focus on studies that use the RCADS-25 (child and parent versions) to measure the constructs it purports. The RCADS-25 is a shortened, brief version of the RCADS-47, which was developed to measure broad anxiety and depression symptoms in children and adolescents as defined by the DSM. Neither the RCADS-25 nor the RCADS-47 were developed to measure other constructs, such as well-being or the lived experience of depression or anxiety, although studies might have extended the instrument's use for these purposes.

Population: We will include all ages, genders, sex, nationalities, languages, countries, cultural backgrounds, and settings of children for the RCADS-25 child version and their parents/caregivers as respondents for the parent version. This includes clinical (physical and mental health) and non-clinical populations. Measurement property evidence syntheses following COSMIN guidelines will be conducted by scopes of interest, defined by distinct populations based on clinically and socio-demographically meaningful categories.

Type of instrument: Child and Parent Versions of the RCADS-25 (including adaptations made into different cultures and languages). Although both the child and parent versions of the RCADS-25 and RCADS-47 will be included in the search and screening phases and mapped for eligibility, Phase B of this review (see section 8) will synthesize only the child and parent versions

of the RCADS-25. Evidence syntheses for eligible RCADS-47 studies (child and parent versions) will be conducted separately as part of a different project.

Measurement properties (based on the COSMIN framework, with adaptations):

1. Content validity
2. Structural validity
3. Internal consistency
4. Cross-cultural validity / measurement invariance
5. Reliability
6. Measurement error
7. Criterion validity
8. Diagnostic and screening accuracy
9. Hypothesis testing for construct validity
10. Responsiveness

Other properties:

11. Feasibility
12. Interpretability
13. Acceptability

Studies: We will include primary studies regardless of their design, as eligibility will be determined by the presence of evaluable evidence on measurement properties rather than by study design labels.

4. Pilot feasibility study

We conducted a preliminary literature search in clinical samples ($\geq 40\%$ of the sample with either depression or an anxiety disorder) to inform the feasibility of the proposed review. This pilot study was conducted by running searches in OVID MEDLINE, Embase, and PsycINFO (from inception until June 13, 2024). The search was conducted for any version of the RCADS (e.g., RCADS-25 or RCADS-47) in all clinical populations with anxiety and/or depressive disorders, yielding 567 records, of which 357 were duplicates. Across this pilot search, a total of 210 records were screened by two independent reviewers at title/abstract level (the results of this pilot study are provided in Appendix A). Of these, 57 (27%) were included in the phase of full-text review, out of which 5 (2%) were deemed eligible for the systematic review. These pilot data demonstrate that there is a sufficient body of the RCADS literature in clinical populations to support the feasibility of the planned systematic review. These pilot studies allowed us to refine the search strings, eligibility criteria, and screening procedures.

5. Eligibility criteria

Studies will be included if they meet the following inclusion criteria:

1. Peer-reviewed articles examining at least one or more measurement properties of the RCADS-25 or RCADS-47 child and/or parent versions. Articles in which measurement

property evaluation is stated as an objective of the paper and report results on measurement properties in the methods or results section will be included if the:

- a. The construct of interest are anxiety and/or depression symptoms, consistent with the original conceptualisation and intended use of the RCADS-25 or RCADS-47 (Ebesutani et al., 2012). If articles do not specify the construct they aim to measure with the RCADS-25 or RCADS-47 (child and/or parent versions), we will assume that the original construct (anxiety and depression symptoms) was measured if consistent with the entire article.
 - b. Measurement properties will be defined according to the COSMIN taxonomy (Mokkink et al., 2010, 2024), with adaptations, as specified in Table 1. Any statistical methods of analysis will be included (e.g., Classic test theory, Item Response theory). Definitions of all measurement properties are provided in Table 1, and further operationalized in Appendix B.
2. There will be no restrictions on population age, sexes/genders, languages, countries, cultural backgrounds, or settings. Clinical (physical and mental health) and non-clinical populations will all be included.
3. Any adaptation of the RCADS-25 child or parent versions will be considered, as long as it maintains the original item content (i.e., 15-item Anxiety Total and 10-item Depression Total scales), such as adaptations to other cultures or populations. Modifications in the wording of items will also be considered if the meaning is not substantially changed. Modified versions of the RCADS will be explicitly recorded and considered during data extraction and synthesis.
4. Any adaptation of the RCADS-47 child or parent versions will be considered, as long as it maintains the original item content (i.e., 37-item Anxiety Total and 10-item Depression Total), such as adaptations to other cultures or populations. Modifications in the wording of items will also be considered if the meaning is not substantially changed. Modified versions of the RCADS will be explicitly recorded and considered during data extraction and synthesis.
5. All self-report formats, including electronic, paper, read-aloud, or interview modalities, of the RCADS-25 or RCADS-47 child version will be eligible for inclusion, as long as the participant is the main respondent for the child version and is free of influence by another individual administering the instrument. Populations with severe impairment will be included if individuals can independently understand, interpret, and respond to the RCADS-25 or RCADS-47.
6. All self-report formats, including electronic, paper, read-aloud, or interview modalities, of the RCADS-25 or RCADS-47 parent version will be eligible for inclusion, as long as the parent/caregiver is the main respondent for the parent version and is free of influence by another individual administering the instrument.
7. We will include primary studies of any design, as evidence of evaluable measurement properties will be used to determine eligibility rather than study design alone. This includes instrument development and validation studies, cross-sectional, longitudinal observational studies, and clinical studies or randomized trials reporting analyses on measurement properties. Qualitative studies with child/youth and parent/caregiver participants will also be included if they report evidence on measurement properties (e.g., content validity,

development paper) or other properties of the tool (e.g., acceptability, feasibility, interpretability).

Table 1. *Definitions of measurement properties and other properties of the measurement tool.*

Measurement properties and other properties of the tools	Definition
1. <i>Content validity</i>	The degree to which the content of a PROM is an adequate reflection of the construct to be measured.
2. <i>Structural validity</i>	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured.
3. <i>Internal consistency</i>	The degree of the interrelatedness among the items.
4. <i>Cross-cultural validity / measurement invariance</i>	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM.
5. <i>Reliability</i>	The proportion of the total variance in the measurements which is due to 'true' ¹ differences between respondents. It indicates an instrument's ability to distinguish between individuals - with various scores and despite measurement error (Mokkink et al., 2023; 2026).
6. <i>Measurement error</i>	The systematic and random error of a respondent's score that is not attributed to true changes in the construct to be measured. It refers to how similar or close the scores are from repeated measurements in an individual who is stable on the outcome of interest (Mokkink et al., 2023; 2026).
7. <i>Criterion validity</i>	The degree to which the scores of a PROM are an adequate reflection of a 'gold standard'.
8. <i>Diagnostic and screening accuracy</i>	The ability of the PROM to correctly classify study participants, as having the target disorder or not. In most cases, that information will be relevant when the PROM is considered for diagnostic reasons (to identify the disorder that most likely is responsible for the signs and symptoms of the patient) or for screening purposes (to identify patients with unrecognised depression).
9. <i>Hypothesis testing for construct validity</i>	The degree to which the scores of a PROM are consistent with hypotheses (regarding relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the PROM validly measures the construct to be measured.
10. <i>Responsiveness</i>	The ability of a PROM to detect change over time in the construct to be measured.
11. <i>Feasibility</i> ²	Ease of application of the PROM in its intended context of use, given constraints such as time or money.

12. *Interpretability*²

The degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations – to a PROM's quantitative scores or change in scores.

13. *Acceptability*²

How respondents perceive the experience of completing the scale, e.g., perceived intrusiveness, burden, potential harms or benefits.

Notes: The definitions of the measurement properties are based on the COSMIN taxonomy, extracted from the COSMIN 2.0 guidelines (Mokkink et al., 2024). The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy will be used where appropriate to supplement the COSMIN guidance (Deeks et al., 2023).

¹The word 'true' must be seen in the context of the CTT, which states that any observation is composed of two components – a true score and error associated with the observation. 'True' is the average score that would be obtained if the scale were given an infinite number of times. It refers only to the consistency of the score, and not to its accuracy (Streiner & Norman, 2008).

²Feasibility, interpretability and acceptability are not considered a measurement property, but important aspects when selecting measurement instruments. Acceptability was added specifically to this project and the definition will be refined according to the literature identified during the systematic review.

The following exclusion criteria will be applied:

1. We will exclude non-peer reviewed materials, such as conference abstracts, posters, dissertations, and other short communications.
2. We will exclude studies in which the RCADS-25 or RCADS-47 (child and/or parent versions) is used solely as a comparator to evaluate another instrument or only as a predictor, grouping variable, or prevalence estimator without any assessment of the instrument's measurement properties.
3. Case studies will be excluded because they will not provide sufficient information or sample size to evaluate measurement properties.

Some studies that do not meet our eligibility criteria will be systematically identified and tagged under predefined categories during both the title and abstract and full-text screening process. Although, these studies fall outside the primary scope of the present review, they will be relevant for parallel projects, methodological insight, or contextual interpretation of the literature.

Studies will be tagged if any of the following criteria are met:

1. Potential Clinician-Reported Outcome Measure (ClinROM) and non-parent or caregiver Observer-Reported Outcome Measure (ObsROM) versions of RCADS-25 and/or RCADS-47 that were used to capture the perspective or experience of clinicians/observers, rather than the respondents (i.e., child or parent/caregiver).
2. Studies that investigated the effects of the RCADS-25 or RCADS-47 administration on individuals, groups, or settings, such as psychological impact, behavioural consequences, or social effects. Such studies aim to measure the consequences of measurement rather than the instrument's measurement properties. They will be integrated in an ancillary scoping review.
3. Studies that provide conceptual, methodological, or implementation-oriented insights into the use of the RCADS-25 or RCADS-47; for example, studies examining barriers and facilitators

to uptake, training strategies for clinicians, or integration into routine care or digital systems. They will be used to guide the website interface of the platform.

4. Studies examining the RCADS-25 or RCADS-47 alongside other measures to generate cross-walks, linking functions, or score equivalence across instruments. The primary focus of these studies is on mathematical comparability across scales rather than on examining the measurement performance of the targeted PROMs itself.
5. Studies that used the RCADS-25 or RCADS-47 to measure a construct beyond anxiety and/or depression will be excluded from the measurement properties synthesis (e.g., well-being, general mental health).
6. Systematic reviews summarizing evidence on measurement properties of the RCADS-25 or RCADS-47, either alone or in comparison with other depression or anxiety measures. We will not include these reviews as primary evidence but will use the reviews to cross-check coverage of primary studies and to identify additional eligible studies that were missed in the original database searches.
7. Studies examining shortened, or modified versions of the RCADS-25 in which the content of the original items has been modified (e.g., RCADS-20, RCADS-11). This includes modifications to response options (e.g., alternative response formats), scoring algorithms (e.g., alternative ways of summing the items into a total score), cut-off thresholds (e.g., exploring different thresholds than the originally validated), instructions for completion (e.g., more developed or specific instructions than the originally provided) or recall period (e.g., experience in the last week instead of the last two weeks).

Search strategy:

To identify relevant studies, we will conduct systematic searches in major bibliographic databases: MEDLINE via PubMed, CINAHL via EBSCO, Embase via Elsevier, Web of Science via Clarivate, and PsycInfo via EBSCO. Grey literature will not be searched. Although grey literature can provide relevant content to increase the comprehensiveness of a review and reduce publication bias, the systematic review aims to map and synthesize peer reviewed studies only to ensure the feasibility and quality of the overall project. We acknowledge that restricting inclusion to peer-reviewed literature may result in the omission of locally produced knowledge, rare language adaptations, and context-specific validation studies that are not indexed in major international bibliographic databases. The searches will be conducted from January 1st, 1995 until January 20th, 2026. As the RCADS-25 is a part of the COLIVE-MH project, the lower bound was chosen based on the data available for the other two scales (i.e., PHQ-9 and WHODAS II), to allow a sufficient margin prior to the instruments' first development papers. An update of the searches will be conducted on July 1st, 2026, and thereafter, updated every 3 months.

The search strategy is structured around two main concept blocks: (1) the RCADS and (2) measurement properties. These blocks will be combined using the Boolean operator AND. The RCADS block is instrument-focused and does not include disorder-related terms. The measurement properties block is based on a validated search filter developed by the COSMIN group to identify studies on measurement properties (Terwee et al., 2009; <https://www.cosmin.nl/tools/pubmed-search-filters/>), which was supplemented with additional

keywords across all measurement properties. Database-specific subject headings (e.g., MeSH, Emtree) and syntax were used where applicable.

The search strategy was initially developed by the review team. It was then reviewed and translated for all bibliographic databases by two information specialists from Université Paris Cité, France. Finally, the full search strategy underwent independent peer review by an external information specialist from the Hospital for Sick Children, Toronto, Canada, in accordance with the PRESS 2015 guidelines (McGowan et al., 2016), prior to execution. The search strategies are provided in Appendix C.

We will also screen the reference lists of included primary studies and relevant systematic reviews to identify additional eligible studies. Citation tracking tools (e.g. Research Rabbit <https://www.researchrabbit.ai/>) will also be used to identify related studies, including forward and backward citations of included articles.

7. Screening procedure

Records will be imported into the Sustainable Knowledge (SK) platform developed by Epistemonikos as part of its end-to-end system for deduplication and reference management (Rada et al., 2020; <https://www.skplatform.org>). The platform will also be used for the screening process. Two independent reviewers will screen all records by title and abstract, with disagreements resolved by a third reviewer. All studies retrieved as full text will be subsequently reviewed by the two reviewers for final inclusion in the systematic review. Disagreements at this stage will be resolved through discussion, and if necessary, involving a third reviewer. The study selection process will be documented using a PRISMA 2020 flow diagram, with the number and reasons for exclusion at the full-text stage provided.

8. Sequential approach for data extraction and synthesis

Data extraction and synthesis will be organized in a workflow of two sequential and complementary phases:

1) Phase A: Living Mapping of the Measurement Evidence Base

In this initial phase we will systematically map the available evidence by identifying all articles that report one or more studies on measurement properties of the RCADS-25 or RCADS-47 child and parent versions (herein referred to as RCADS) or report on the interpretability of the RCADS (child or parent versions) and classify their broad characteristics (including study features, instrument characteristics (e.g., language version used), participant characteristics, and the measurement properties assessed).

The search and screening phase will include the child and parent versions of the RCADS-25 or RCADS-47. Both the RCADS-25 and RCADS-47 will be screened for eligibility and mapped accordingly. For the purpose of Phase B of this review, we will however synthesize the child and parent versions of the RCADS-25 solely, while the evidence syntheses for eligible studies of the child and parent versions of the RCADS-47 will be completed for a separate project.

2) Phase B: Living Evidence Syntheses of the Measurement Properties

In this phase we will conduct COSMIN-guided systematic reviews on scopes of interest for the RCADS-25 (child and parent versions separately), such as specific populations or intended instrument uses. The reviews will synthesize and rate the quality of the results, assess risk of bias, and grade the certainty of evidence.

The outputs of Phase A and Phase B will be disseminated in an open-access platform with summaries, visualisations, and structured syntheses of measurement properties, as well as (child and/or parent versions) across contexts. The project will expose evidence gaps across populations, settings, and measurement properties, informing areas where future primary research is needed. In addition, we will identify potential areas for improvement of the RCADS-25 (child and/or parent versions).

9. Phase A: Living mapping of the evidence

Data extraction. All studies meeting the inclusion criteria after full-text screening will undergo structured data extraction within the Epistemonikos end-to-end system. Two reviewers will independently extract data. All the reviewers involved in the screening will train on 10% of the included studies. The codebook for data extraction for the Living Mapping phase is provided in Appendix D. For each study, data will be collected across eight broad categories:

- A. **Study metadata:** Study ID, year of publication, journal, DOI, etc.
- B. **Study characteristics:** Design of the study, inclusion criteria with threshold, etc
- C. **Characteristics of the RCADS-25 and RCADS-47:** Exact version of the instrument used (child, parent), administration mode (self-administered, read-aloud, etc.), format (Paper/pencil, online, etc.), different languages and cultural adaptations (e.g., translated to French), or other adaptations.
- D. **Purpose of use** for which the properties of the RCADS-25 and RCADS-47 (child and parent versions) are being examined (outcome assessment, screening/diagnostic).
- E. **Participant characteristics:** Sociodemographic (age, gender/sex, and country, specific demographics such as prisoners, migrants, etc) and clinical characteristics (e.g., general population, physical illness, mental health condition) etc.
- F. **Measurement properties:** At this stage, we will only classify which measurement properties are examined in each article. Measurement properties will be classified based on the analyses and statistics reported in the primary studies, following COSMIN guidance. Definitions of all measurement properties are listed in Table 1, and examples of commonly reported analyses and indicators are operationalized in Appendix B.
- G. **Other properties of the tool:** Studies reporting evidence on the feasibility, interpretability and/or acceptability of the RCADS-25 and RCADS-47 (child and parent versions) will be also classified accordingly.

- H. **Involvement of contributors with lived experiences:** We will extract information on whether individuals with lived experiences actively contributed to each included study, beyond the role of participants.

The data extraction form will be pilot tested on a subset of random included studies to ensure completeness and consistency. In the context of a living systematic review, both the extraction form and the dataset will be updated on an ongoing basis as new evidence emerges.

Data synthesis. The data extraction of this phase will allow us to map the available evidence and to inform the organization of Phase B. The data will be analysed with descriptive statistics, and all outputs will be displayed in the open-access platform.

10. Phase B: Living evidence syntheses of the RCADS-25 properties

Phase B will synthesize evidence on the child and parent versions of the RCADS-25's measurement properties separately following COSMIN guidance. We will first extract and assess RCADS-25 characteristics from its original development studies to define their measurement scope and support the later assessment of content validity. Based on the results of the living evidence mapping conducted in Phase A, we will define clinically and sociodemographically meaningful scopes of interest (for instance properties of the RCADS-25 across different age groups) and conduct separate COSMIN-guided systematic reviews within each scope.

The selection and sequencing of these scopes will be informed by the mapped evidence base and by input from project collaborators, and external advisory boards, including representatives of people with lived experiences. All systematic reviews will be conducted using harmonized methods to facilitate comparison and integration of findings across scopes.

For each scope of interest, we will extract data on all reported measurement properties and other properties (e.g., interpretability) of the RCADS-25 child and parent versions. We will assess the methodological quality of included studies using the COSMIN risk of bias checklist, evaluate results following the COSMIN criteria for good measurement properties, summarize findings (including quantitative pooling where appropriate), and assess the certainty of evidence using the GRADE approach per measurement property. The present project also aims to adapt and refine the COSMIN risk of bias checklist and the GRADE criteria for certainty of synthesised evidence of measurement properties, with details available in other research protocols.

Further details on the procedures for the evidence syntheses are reported in the Master Protocol of the COLIVE-MH project.

11. Living systematic review methodology

This review is designed as a living systematic review. Following the initial search, literature searches will be updated after 6 months to allow enough time for the initial screening and data extraction, and then at subsequent 3-month intervals. However, frequency of updates may be adapted based on the amount of evidence and the available resources. At each update, newly identified records will undergo title and abstract screening and full-text assessment using the

same eligibility criteria and standardized methodological procedures as those applied in the initial review. The newly identified eligible studies will be included in the data extraction syntheses of Phase A and B. Every update will be disseminated in partnership through the Epistemonikos platform for Interactive Evidence Synthesis (<https://www.skplatform.org>).

The review is planned to remain in living mode for a period of 2 years starting from the date of the first search. Retirement from the living mode will be considered based on prespecified criteria, including stability of conclusions across updates (e.g., reaching High GRADE), or reduced relevance of the review question. If none of these criteria are met at the end of the planned period, the need for continuing the updates will be reassessed.

12. Ethics and dissemination

As this study synthesizes existing literature, no ethical approval will be required. The protocol will be prospectively registered on PROSPERO (<https://www.crd.york.ac.uk/prospero/>), and on Zenodo (<https://zenodo.org>) and all protocol deviations will be transparently reported in the Supplementary Materials of the respective peer-reviewed publications issued from this living systematic review. Findings will be disseminated through peer-reviewed open-access publications, national and international conference presentations. Results will be hosted on an open-access living evidence platform which development will be described elsewhere.

13. Lived experience engagement

We consider the engagement of people with lived experiences of mental health conditions as key to this project since the RCADS-25 child and parent versions and the three other measurement tools intend to measure mental health symptoms through self-reported evaluation of symptoms. As of February 9th, 2026, the engagement plan is still under development with people with lived experiences from different countries and with different backgrounds. This protocol has been developed with co-researchers with lived experiences and background in psychology and is under review by three lived experts with background in anthropology. Their feedback will be implemented in the next version. The next update of this protocol will display the precise procedures of engagement at each stage of the LSR.

References

- Akl, E. A., Khabisa, J., Iannizzi, C., Piechotta, V., Kahale, L. A., Barker, J. M., McKenzie, J. E., Page, M. J., & Skoetz, N. (2024). Extension of the PRISMA 2020 statement for living systematic reviews (PRISMA-LSR): Checklist and explanation. *BMJ*, 387, e079183. <https://doi.org/10.1136/bmj-2024-079183>
- Bear, H. A., Edbrooke-Childs, J., Norton, S., Krause, K. R., & Wolpert, M. (2020). Systematic Review and Meta-analysis: Outcomes of Routine Specialist Mental Health Care for Young People With Depression and/or Anxiety. *Journal of the American Academy of Child & Adolescent Psychiatry*, 59(7), 810–841. <https://doi.org/10.1016/j.jaac.2019.12.002>
- Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, 38(8), 835–855. [https://doi.org/10.1016/S0005-7967\(99\)00130-8](https://doi.org/10.1016/S0005-7967(99)00130-8)
- Ebesutani, C., Korathu-Larson, P., Nakamura, B. J., Higa-McMillan, C., & Chorpita, B. (2017). The Revised Child Anxiety and Depression Scale 25–Parent Version: Scale Development and Validation in a School-Based and Clinical Sample. *Assessment*, 24(6), 712–728. <https://doi.org/10.1177/1073191115627012>
- Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., Higa-McMillan, C., & Weisz, J. R. (2012). The Revised Child Anxiety and Depression Scale-Short Version:

- Scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychological Assessment*, 24(4), 833–845. <https://doi.org/10.1037/a0027283>
- Elsman, E. B. M., Mokkink, L. B., Terwee, C. B., Beaton, D., Gagnier, J. J., Tricco, A. C., Baba, A., Butcher, N. J., Smith, M., Hofstetter, C., Aiyegbusi, O. L., Berardi, A., Farmer, J., Haywood, K. L., Krause, K. R., Markham, S., Mayo-Wilson, E., Mehdipour, A., Ricketts, J., ... Offringa, M. (2024). Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *Quality of Life Research*, 33(8), 2029–2046. <https://doi.org/10.1007/s11136-024-03634-y>
- Farber, G. K., Gage, S., Kemmer, D., & White, R. (2023). Common measures in mental health: A joint initiative by funders and journals. *The Lancet Psychiatry*, 10(6), 465–470. [https://doi.org/10.1016/S2215-0366\(23\)00139-6](https://doi.org/10.1016/S2215-0366(23)00139-6)
- McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of Clinical Epidemiology*, 75, 40–46. <https://doi.org/10.1016/j.jclinepi.2016.01.021>
- Mew, E. J., Monsour, A., Saeed, L., Santos, L., Patel, S., Courtney, D. B., Watson, P. N., Szatmari, P., Offringa, M., Monga, S., & Butcher, N. J. (2020). Systematic scoping review identifies heterogeneity in outcomes measured in adolescent depression clinical trials. *Journal of Clinical Epidemiology*, 126, 71–79. <https://doi.org/10.1016/j.jclinepi.2020.06.013>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1. <https://doi.org/10.1186/2046-4053-4-1>
- Mokkink, L. B., Elsman, E. B. M., & Terwee, C. B. (2024). COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Quality of Life Research*, 33(11), 2929–2939. <https://doi.org/10.1007/s11136-024-03761-6>
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & De Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., Knol, D. L., Bouter, L. M., & De Vet, H. C. W. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research*, 18(3), 313–333. <https://doi.org/10.1007/s11136-009-9451-9>
- Monga, S., Andrei, S., Quinn, R. C., Khudiakova, V., Desai, R., Srirangan, A., Patel, S., Szatmari, P., Butcher, N. J., Krause, K. R., Courtney, D. B., Offringa, M., & Elsman, E. B. M. (2025). Systematic Review: Measurement Properties of Patient-Reported Outcome Measures Used to Measure Depression Symptom Severity in Adolescents With Depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, 64(2), 198–225. <https://doi.org/10.1016/j.jaac.2024.06.011>
- Rada, G., Pérez, D., Araya-Quintanilla, F., Ávila, C., Bravo-Soto, G., Bravo-Jeria, R., Cánepa, A., Capurro, D., Castro-Gutiérrez, V., Contreras, V., Edwards, J., Faúndez, J., Garrido,

- D., Jiménez, M., Llovet, V., Lobos, D., Madrid, F., Morel-Marambio, M., Mendoza, A., ... Zilleruelo-Ramos, R. (2020). Epistemonikos: A comprehensive database of systematic reviews for health decision-making. *BMC Medical Research Methodology*, 20(1), 286. <https://doi.org/10.1186/s12874-020-01157-x>
- Schueler, S., Schuetz, G. M., & Dewey, M. (2012). The Revised QUADAS-2 Tool. *Annals of Internal Medicine*, 156(4), 323. <https://doi.org/10.7326/0003-4819-156-4-201202210-00018>
- Staniszewska, S., Brett, J., Simera, I., Seers, K., Mockford, C., Goodlad, S., Altman, D. G., Moher, D., Barber, R., Denegri, S., Entwistle, A., Littlejohns, P., Morris, C., Suleman, R., Thomas, V., & Tysall, C. (2017). GRIPP2 reporting checklists: Tools to improve reporting of patient and public involvement in research. *BMJ*, j3453. <https://doi.org/10.1136/bmj.j3453>
- Terwee, C. B., Jansma, E. P., Riphagen, I. I., & De Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, 18(8), 1115–1123. <https://doi.org/10.1007/s11136-009-9528-5>

Appendix A: Pilot Feasibility study for the RCADS (child and parent versions)

Purpose of the pilot study:

To refine the methodology of the living systematic review, we conducted a pilot study focused on identifying the most effective way to screen and extract relevant information and evidence on the measurement properties of Revised Child Anxiety and Depression Scale (RCADS) 25-item or 47-item child and parent versions . The pilot was designed to assess the feasibility of our search strategy, screening workflow, inclusion/exclusion criteria and test data extraction procedures, guided by COSMIN recommendations.

Scoping work and development of search strategy:

We began by reviewing existing systematic reviews and studies on the measurement properties using the RCADS-25 and RCADS-47 child and parent versions to understand how authors structured their searches, developed search terms, and reported measurement properties, including how they applied frameworks such as COSMIN. This scoping work informed the development of our initial COSMIN-informed search strategy for RCADS-25 and RCADS-47 in clinical populations with anxiety and/or depressive disorders.

The following databases were searched from inception until June 13, 2024:

- a. Ovid MEDLINE
- b. PsycINFO
- c. Embase

Attached below is the final search string developed and utilised for the pilot study (search string provided for Ovid MEDLINE as an example):

Ovid MEDLINE

#	Searches	Results
1	exp Depression/	158798
2	Depression.tw,kw,kf.	454155
3	exp Depressive Disorder, Major/	40619
4	(depressive disorder* or clinical depression* or unipolar depression* or major depression* or depressive disease* or depressive episode* or depressive illness* or depressive state* or depressive syndrome* or endogenous	92094

	depression* or major depressive episode* or mental depression* or neurotic depression*).tw,kf.	
5	exp Anxiety Disorders/	94195
6	exp Agoraphobia/	2687
7	exp Anxiety, Separation/	2203
8	exp Neurotic Disorders/	18004
9	exp Panic Disorder/	7357
10	exp Phobic Disorders/	14460
11	exp Phobia, Social/	1318
12	exp Mutism/	1177
13	(anxiety disorder? or anxiety neuros?s or neurotic anxiety state? or generali#ed anxiety or panic disorder? or panic attack? or selective mutism* or Mutism? or Elective Mutism? or Voluntary Mutism? or Kussmaul Aphasia? or "Kussmaul's Aphasia" or "Conversion Mutism?" or Organic Mutism? or "social* anxi*" or separation anxiety* or neuros#s or neurotic disorder? or psychoneuros#s or psycho-neuros#s or phobi? or claustrophobi? or hypervigilan* or agoraphobi? or Crowd* Phobi? or Open Spaces Phobia or claustrophobi? or Scolionophobi? or School Phobi? or "Social Evaluation Phobia?" or Sociophobi?).mp.	118641
14	or/1-13	608061
15	exp adolescent/ or exp child/ or *pediatrics/ or (childhood adj2 (disease* or ill* or condition* or syndrom*)).ti,ab,kf. or (teen* or teen-age* or adolescen* or child* or juvenil* or p?ediatric* or paediatric* or puber* or pubescen* or teen* or toddler? or underage? or under-age? or youth*).ti,ab,kf. or (pediatric* or paediatric* or child* or adolescen* or young).jn,jw.	4426866
16	((Revised Child Anxiety and Depression Scales) or RCADS or (Revised Child Anxiety and Depression Scales Depression subscale) or RCADS depression subscale or (Revised Child Anxiety and Depression Scales Anxiety subscale) or RCADS anxiety subscale).mp.	149
17	(instrumentation or methods).sh. or (Validation Studies or Comparative Study).pt. or exp Psychometrics/ or psychometr*.ti,ab. or (clinimetr* or clinometr*).tw. or exp "Outcome Assessment (Health Care)"/ or outcome assessment.ti,ab. or outcome measure*.tw. or exp Observer Variation/ or observer variation.ti,ab. or exp Health Status Indicators/ or exp "Reproducibility of Results"/ or reproducib*.ti,ab. or exp Discriminant Analysis/ or (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or "internal consistency").ti,ab. or (cronbach* adj2 (alpha or alphas)).ti,ab. or (item and (correlation* or selection* or reduction*)).ti,ab. or (agreement or precision or imprecision or "precise values" or test-retest).ti,ab. or (test and retest).ti,ab. or (reliab* and (test or retest)).ti,ab. or (stability or interrater or inter-rater or	9347957

	intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intra-observer or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or inter-assay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappa's or kappas or repeatab*).ti,ab. or ((replicab* or repeated) and (measure? or finding? or result or results or test or tests)).ti,ab. or (generaliz*a* or concordance).ti,ab. or (intraclass and correlation*).ti,ab. or (discriminative or "known group" or factor analys?s or dimension* or subscale*).ti,ab. or ((multitrait* or multi-trait*) and scaling and analys?s).ti,ab. or (item discriminant or interscale correlation* or error or errors or "individual variability").ti,ab. or (variability and (analys?s or values)).ti,ab. or (uncertainty and (measurement or measuring)).ti,ab. or ("standard error" adj measur*).ti,ab. or (sensitiv* or responsive*).ti,ab. or ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change? or difference)).ti,ab. or (small* and (real or detectable) and (change? or difference)).ti,ab. or (meaningful change or "ceiling effect" or "floor effect" or "Item response model" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross-cultural equivalence").ti,ab.	
18	14 and 15 and 16	143
19	17 and 18	92
20	limit 19 to english language	91
21	remove duplicates from 20	91

Screening process:

A PRISMA flow diagram (Figure 1), summarizes the screening process below. The inclusion criteria during both the title/abstract and full-text screening included:

1. Studies available in full text and published in English.
2. Studies that evaluated at least one or more of the nine measurement properties as detailed in the COSMIN taxonomy and/or detailed the development of the RCADS-25 and RCADS-47, and/or the evaluation of the interpretability of the RCADS-25 and RCADS-47. The nine properties as per the COSMIN taxonomy include content validity, structural validity, internal consistency, cross cultural /measurement invariance, test re-test reliability, measurement error, criterion validity, hypothesis testing for construct validity (including convergent validity and discriminative validity), or responsiveness.
3. The measurement properties must have been assessed within a reported clinical sample or subgroup of youth (ages 4 to 24 years) with at least 40% of the population having either an anxiety disorder diagnosis or a diagnosis of major depressive disorder (MDD).

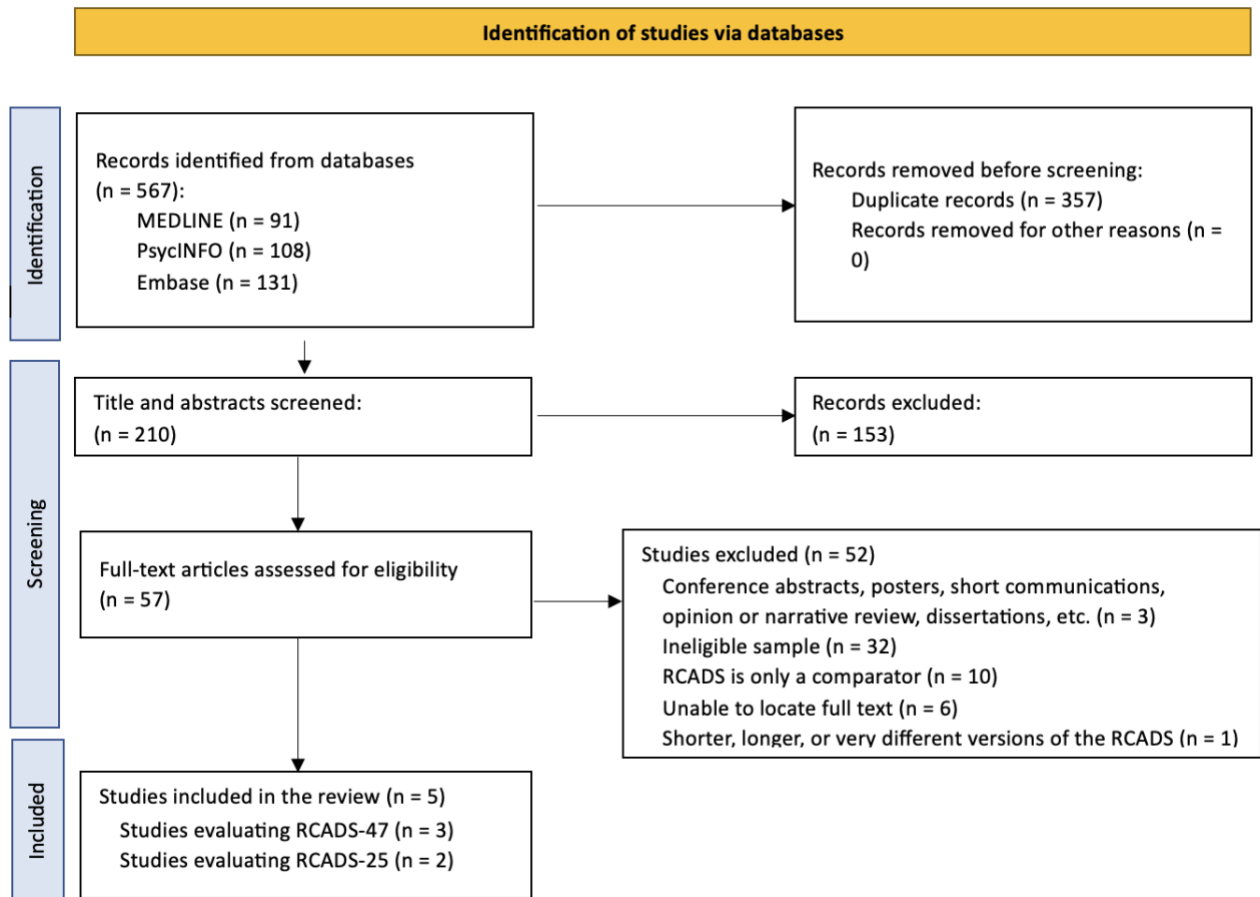
Our exclusion criteria, during both the title/abstract and full-text screening included:

1. Non-peer reviewed publications (e.g. dissertations, conference abstracts).
2. Studies that did not evaluate any measurement properties or evaluated or described any interpretability or feasibility aspects of the RCADS-25 or RCADS-47.
3. Studies evaluating non-standard versions of the RCADS-25 or RCADS-47 (e.g. RCADS-11).
4. Studies that utilised the RCADS-25 or RCADS-47 as a comparator instrument for the validation of other measures of depression/mental health.
5. Studies not published in English.
6. Studies that do not include a population of youths aged 4 to 24 years or reports on a subgroup analysis of participants aged 4 to 24 years (mean or median or subgroup).
7. Studies that do not include a population comprised of children and/or youth with a clinical diagnosis of anxiety disorder(s) (40% of sample) or reports on a subgroup analysis of youth with an AD(s).

Note: Systematic reviews were not included for our study but tagged to provide an overview of the existing literature and utilise them as a base for additional studies that could otherwise be undetectable by our search.

Figure 1:

PRISMA flowchart representing screening process for pilot study on the RCADS-



Data extraction:

Extracted information included general study and participant characteristics, such as sample size (number of child participants and number of parent participants), population demographics, and country where the study was conducted, population gender, language RCADS-25 or RCADS-47 was administered, RCADS-25 or RCADS-47 version (child or parent) and administration format, and measurement properties assessed in the study. The table below (Table 1) provides a summary of the data extracted during our pilot study.

Table 1

Summary of the included studies.

Study	Version assessed	Measurement properties assessed	Number of child participants	Gender (% female)	Demographics	Number of parent participants	Country	Language OMI was administered
Ebesutani (2010)	RCADS-Parent 47 Version	Structural validity, internal consistency, reliability, criterion	490	32.2	Youth Ethnicity: 39.0% Multiethnic 34.7% White	490	United States of America	English

		validity, hypothesis testing for construct validity			7.3% African American 5.7% Asian American 5.1% Latino/Hispanic 5.5% Other 2.7% Missing			
Ebesutani (2012)	RCADS Child Version	Structural validity, internal consistency, criterion validity, hypotheses testing for construct validity	303	32.7	Youth Ethnicity: 48.5% White 27.1% Multiethnic 10.6% African American 7.3% Latino/Hispanic 3.6% Asian American 3.0% Other	303	United States of America	English
Gormez (2017)	Turkish RCADS-47 Child Version	Structural validity, internal consistency, reliability, criterion validity, hypotheses testing for construct validity	483	39.3	Not reported	Not reported	Turkey	Turkish
Gormez (2017)	Turkish RCADS-47 Parent Version	Structural validity, internal consistency, reliability, criterion validity, hypotheses testing for construct validity	483	39.3	Two-parent home: 90.7% Monthly family income (Turkish Lira): 15.1% 0–1000 31.2% 1000–2000 24.8% 2000–3000 12.4% 3000–5000 6.2% 5000 and over 10.4% Missing	Not reported	Turkey	Turkish
Ebesutani (2017)	RCADS-25 Parent Version	Structural validity, internal consistency, reliability, criterion validity, hypotheses testing for	433	32.1	Youth Ethnicity: 60.0% Multiethnic 12.5% Asian 10.9% White 1.6% Black 9.5% Other 5.5% Not reported	397	United States of America	English

		construct validity						
--	--	--------------------	--	--	--	--	--	--

Refinement of methods

Based on insights from this pilot work, refinements to the search strategy and screening procedures were made in collaboration with the review team, COSMIN search filters, and research librarians to improve precision while maintaining comprehensive coverage. The finalized search strategy will be applied across all selected databases in the full mapping phase.

References of included studies

- Ebesutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., Weisz, J. R., & The Research Network on Youth Mental Health. (2010). A Psychometric Analysis of the Revised Child Anxiety and Depression Scale—Parent Version in a Clinical Sample. *Journal of Abnormal Child Psychology*, 38(2), 249–260. <https://doi.org/10.1007/s10802-009-9363-8>
- Ebesutani, C., Korathu-Larson, P., Nakamura, B. J., Higa-McMillan, C., & Chorpita, B. (2017). The Revised Child Anxiety and Depression Scale 25—Parent Version: Scale Development and Validation in a School-Based and Clinical Sample. *Assessment*, 24(6), 712–728. <https://doi.org/10.1177/1073191115627012>
- Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., Higa-McMillan, C., & Weisz, J. R. (2012). The Revised Child Anxiety and Depression Scale-Short Version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychological Assessment*, 24(4), 833–845. <https://doi.org/10.1037/a0027283>
- Gormez, V., Kilincaslan, A., Ebesutani, C., Orengul, A. C., Kaya, I., Ceri, V., Nasiroglu, S., Filiz, M., & Chorpita, B. F. (2017). Psychometric Properties of the Parent Version of the Revised Child Anxiety and Depression Scale in a Clinical Sample of Turkish Children and Adolescents. *Child Psychiatry & Human Development*, 48(6), 922–933. <https://doi.org/10.1007/s10578-017-0716-1>
- Gormez, V., Kılınçaslan, A., Orengul, A. C., Ebesutani, C., Kaya, I., Ceri, V., Nasıroğlu, S., Filiz, M., & Chorpita, B. (2017). Psychometric properties of the Turkish version of the Revised Child Anxiety and Depression Scale – Child Version in a clinical sample. *Psychiatry and Clinical Psychopharmacology*, 27(1), 84–92. <https://doi.org/10.1080/24750573.2017.1297494>

Appendix B. Operationalization of the Measurement Properties across the Project

The table below provides operational definitions of each measurement property considered in COLIVE-MH, together with practical guidance for reviewers on how to identify relevant evidence in primary studies. For each property, we specify its conceptual definition and indicate the types of analyses, statistics, or study features that reviewers should look for (but not limited to) when extracting and classifying evidence on measurement properties. We also included other properties of the measurement tools that are not measurement properties but constitute important aspects when using a tool (i.e., feasibility, interpretability, acceptability). These definitions do not take into account the context of use (population, purpose) of the measure and are theoretical. Nevertheless, we acknowledge that measurement properties are the product of an interaction between a measurement tool and a population, meaning that these properties might vary across populations. For instance, convergent validity might be correct in a sample of people with diagnosed depression but not in a sample of people with no mental health disorders. The definitions and operationalizations are based on the COSMIN guidelines which is the most comprehensive framework we identified (Mokkink, Elsmans, & Terwee, 2024). However, we acknowledge that measurement science is a dynamic field and that alternative definitions, methods and further progress can be made to conceptualize, assess and interpret measurement properties.

Measurement (and other related) properties	Definitions and Operationalization
1. Content validity	Definition: Content validity is the degree to which the content of an instrument adequately reflects the construct to be measured. It comprises three aspects: relevance (whether items are pertinent to the construct and population), comprehensiveness (whether all key aspects of the construct are covered), and comprehensibility (whether items, instructions, response options, and recall period are understood as intended). Typical indicators: Content validity evidence may come from concept elicitation studies (e.g., qualitative interviews or focus groups) assessing whether all key concepts are represented; from pilot or cognitive interviewing studies assessing comprehensibility of items, instructions, response options, and recall period; or from separate content validity studies in which people with lived experiences or professionals judge relevance, comprehensiveness, and comprehensibility. Analyses may include e.g., content analysis, framework analysis, deductive analysis, or grounded theory.

2. Structural validity	Definition: Structural validity refers to the degree to which the scores of a PROM adequately reflect the dimensionality of the construct to be measured, that is, whether the items group into the expected factor structure or latent model. Typical indicators: Structural validity is commonly evaluated during the development of the measures using principal component analysis (PCA) or exploratory factor analysis (EFA). Structural validity is further confirmed with confirmatory factor analysis (CFA). Reported results may include the number of factors, item–factor loadings, explained variance, eigenvalues, model fit indices (e.g., CFI, TLI, RMSEA, SRMR), and tests of model assumptions such as unidimensionality, local independence, monotonicity, or item fit statistics. Item response theory (IRT) or Rasch models can also be used to examine the items in a scale.
3. Internal consistency	Definition: Internal consistency refers to the degree of interrelatedness among items in a scale. It reflects whether items consistently measure the same underlying construct, assuming unidimensionality. Typical indicators: Internal consistency is typically quantified using Cronbach’s alpha, McDonald’s Omega, or IRT/Rasch-based reliability indices (e.g., standard error of theta). Interpretation of these statistics requires evidence that the items form a unidimensional scale, which should be established through prior assessment of structural validity.
4. Cross-cultural validity and measurement invariance	Definition: Cross-cultural validity (or measurement invariance) refers to the extent to which PROM items perform equivalently across different groups, such as languages, cultures, genders, age groups, or clinical populations. It assesses whether individuals with the same underlying level of the construct respond similarly to the items, regardless of group membership. Typical indicators: Cross-cultural validity is commonly evaluated using multi-group factor analysis, including tests of configural, metric, and scalar invariance, or through differential item functioning (DIF) analyses. Studies may also examine longitudinal measurement invariance to assess score comparability over time. Comparisons typically involve culturally or linguistically distinct groups, and may also include other relevant subgroups (e.g., gender, ages etc.).

5. Reliability	Definition: Reliability refers to the proportion of total variance in scores that is due to true differences between individuals. It reflects the ability of an instrument to distinguish between respondents. Typical indicators: Reliability is evaluated using repeated measurements in participants who are expected/assumed to remain stable on the construct. Measurements should be conducted under identical conditions, including the same PROM version, administration mode, response format, and setting, except for the specific source of variation being examined. Most commonly, reliability is assessed over time (test–retest reliability), but studies may also examine consistency across administration modes (e.g., paper vs digital), response formats (e.g., Likert vs VAS), or administrators (e.g., self-report vs interviewer-assisted/read aloud). Reported statistics typically include intraclass correlation coefficients (ICC) or kappa values, with specification of the ICC model when applicable. Some studies report Pearson or Spearman correlations, although these are less appropriate for assessing reliability.
6. Measurement error	Definition: Measurement error refers to the systematic and random error in a PROM score that is not attributable to true change in the construct being measured. It refers to the precision of the score. Measurement error is sometimes referred to as agreement and is evaluated using repeated measurements in participants who are assumed to be stable on the construct. The same data collected for the assessment of reliability can be used to evaluate measurement error. Typical indicators: Measurement error is assessed using repeated measurements in stable participants, often drawing on the same study designs used for reliability analyses. Reported statistics may include the standard error of measurement (SEM), smallest detectable change (SDC), limits of agreement (LoA), or percentage agreement for categorical scores. These indices quantify the amount of error around an observed score or change score. Kappa statistics are not measures of measurement error but of reliability.
7. Criterion validity	Definition: Criterion validity refers to the degree to which PROM scores are an adequate reflection of a reference standard (gold standard). Definition by the review team: Upon discussion within the review team, we agreed to define as gold standard the reporting of symptom counts based on clinical diagnostic interviews such as the Anxiety Disorders Interview Schedule for DSM 5 (ADIS-5) or for the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS). These diagnostic

	interviews assess for diagnoses of various anxiety disorders or mood disorders and as such include symptoms assessed by the RCADS-25. Symptom counts based on these diagnostic interviews are the most reasonable “gold standard”. They are comparable to measurement of self-reported symptom severity of anxiety and/or depression. Additionally, a clinical diagnosis by a trained professional (e.g., psychiatrist or psychologist) who can make diagnoses of an anxiety disorder or depressive disorder would also be considered as a “gold standard.” Psychiatric diagnoses are inherently clinical, requiring clinician-patient interaction and consideration of functional impairment, symptom duration, and exclusion of medical mimics. The RCADS-25 cut-off thresholds should not be used independently for diagnosing any of the anxiety disorders (e.g., generalized anxiety disorder) or major depression disorder, as the RCADS-25 omits these essential diagnostic components (Ebesutani et al., 2012; Ebesutani et al., 2017). Instead, the RCADS-25 serves as a rating of a child/adolescent’s anxiety or depressive symptoms alerting the need for further comprehensive clinical evaluation. Consequently, studies that provide a formal diagnosis based upon the RCADS-25 alone does not constitute a suitable gold standard for RCADS-25 criterion validity. The performance of thresholds of the RCADS-25 for screening or as a proxy of diagnosis will be examined on a separate measurement property (8. Diagnostic accuracy).
8. Diagnostic and screening accuracy	<u>Definition by the review team</u> The review team will consider the accuracy of the RCADS-25 when evaluated for both diagnostic (diagnosis of the disorder causing the symptoms in clinical population) and screening (identifying unrecognised disorder in non-clinical population) purposes. Diagnostic and screening accuracy can be defined as the ability of a PROM to correctly classify study participants, based on the results, as having the disorder or not (Deeks & Bossuyt, 2023, Chapter 2). <u>Typical indicators</u> We will include studies examining the accuracy of the RCADS-25 against a binary diagnostic outcome, defined using a structured or semi-structured clinical interview based on established diagnostic criteria (e.g., DSM or ICD). Typical indicators of diagnostic accuracy are: sensitivity, specificity, positive and negative predictive values, likelihood ratios, diagnostic odds ratios, area under the receiver operating characteristic curve (AUC), and 2×2 classification tables. Studies may assess one or multiple cut-off values to examine classification performance across thresholds, often using ROC analyses.
9. Hypothesis testing for construct validity	Definition: Hypothesis testing for construct validity evaluates whether PROM scores behave as expected in relation to other measures or groups, based on a priori hypotheses defined by the review team. Hypotheses formulated by the review team:

	In the section of Hypothesis testing for construct validity, the RCADS-25 is assessed within the purpose of evaluating symptom severity, and hypotheses are formulated accordingly. <u>Using the RCADS as a measure of symptom severity</u> Phase A will map the studies reporting on construct validity, and this will provide an overview of the comparator instruments used. We expect that for convergent validity, common comparators will be instruments that measure anxiety symptom severity, for example, the Screen for Child Related Anxiety Disorders (SCARED), or the Spence Children's Anxiety Scale (SCAS); while for and for depression symptom severity, common comparators will be, for example, the Mood and Feelings Questionnaire (MFQ) or the Children's Depression Inventor (CDI). Hypothesis testing for construct validity will also be completed for comparisons between the child- and parent-reports of the RCADS-25. Further hypotheses for testing construct validity will be specified after the mapping step based on the available comparators. We will consider doing further evidence synthesis (Phase B) on convergent, divergent, and discriminant validity depending on research community needs and available data, discussed with the lived experiences team members and clinicians. Usual indicators reported in the included studies will be correlations between the RCADS-25 total scores and the available comparator measures.
10. Responsiveness	Definition: Responsiveness refers to the ability of a PROM to detect change over time in the construct to be measured. It concerns the accuracy of a change score in relation to a change in the construct measured. Typical indicators: Responsiveness is evaluated by testing <i>a priori</i> hypotheses about expected patterns of change, including: ● Associations between change scores on the PROM and change scores on comparator instruments measuring the same or related constructs. ● Differences in change between groups expected to differ over time (e.g., intervention vs control groups, responders vs non-responders). ● Expected effect sizes following interventions with known or established effects. Hypotheses should specify the expected direction and, where possible, the expected magnitude of change or association. Reported results may include correlations between change scores, mean differences in change, or effect sizes. Statistical significance alone is not sufficient; emphasis is placed on whether observed changes align with predefined hypotheses. Measures such as Guyatt's responsiveness ratio or MIC/MID are not used to evaluate responsiveness itself, these are measures of interpretability.

	Hypotheses formulated by the review team: <u>RCADS-25</u> Hypotheses are based on the expected direction and magnitude of change scores in the RCADS-25 compared to change scores in other instruments: We expect moderate-to-strong correlations between change scores on the RCADS-25 and change scores on established measurements of anxiety or depression as a diagnosis or as a continuous measure of symptom severity). We expect that the comparator instruments most commonly found will be the same as those for construct validity: • RCADS-25 used for screening or diagnostic purposes: Diagnostic reference standard (e.g., with ADIS-5 or K-SADS or DSM/ICD-based diagnosis). • RCADS-25 for assessing symptom severity: other validated measures (PROMs and ClinROMs) commonly used in clinical research for evaluating symptom severity, such as the SCARED, SCAS for anxiety or the MFQ or CDI for depression. Phase A will map the studies reporting on responsiveness, and this will provide an overview of the comparator instruments used. Definitive hypotheses for responsiveness will be specified after the mapping step based on the available comparators. As for Hypothesis testing for construct validity, we will synthesize evidence for responsiveness (Phase B) for the most commonly used comparators, based on the needs of the research community and upon discussion with the members of the COLIVE-MH team.
11. Feasibility	Definition: Feasibility refers to the ease of administering the PROM in its intended context, considering practical constraints such as time, burden, cost, and accessibility. Typical indicators: Evidence on feasibility may be reported in studies describing the practical use of the PROM, including information on completion time, number of items, mode of administration, availability of translations, licensing or permission requirements, and implementation-related constraints. Some studies may also report indicators of acceptability or uptake.
12. Interpretability	Definition: Interpretability refers to the degree to which meaningful qualitative information can be assigned to PROM scores or change scores. Interpretability does not evaluate measurement quality itself but provides context for understanding what specific scores or changes represent. Typical indicators: Studies may report the distribution of scores, including floor and ceiling effects, reference values, or cut-off

	scores indicating severity levels (e.g., mild, moderate, severe). For change scores, studies may report minimal important change (MIC) or minimal important difference (MID), representing the smallest within-person change considered important. Although MIC/MID values are primarily used to interpret change, they are also necessary when evaluating measurement error.
13. Acceptability	Definition: How respondents perceive the experience of completing the scale, e.g., perceived intrusiveness, burden, potential harms or benefits. Typical indicators: Time of completion, number of missing responses, satisfaction with the questionnaire, qualitative reports on the experience of using the scale in research or clinical settings, etc.

Appendix C: Search strings

Block #1: RCADS keywords

("RCADS"[TIAB] OR "revised child* anxiety and depression scale"[TIAB])

Block #2: COSMIN filter

AND ("instrumentation"[sh] OR "methods"[sh] OR "Validation Study"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR "psychometr*[tiab] OR "clinimetr*[tw] OR "clinometr*[tw] OR "Outcome Assessment, Health Care"[Mesh] OR "outcome assessment"[tiab] OR "outcome measure*[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR "reproducib*[tiab] OR "discriminant analysis"[MeSH] OR "reliab*[tiab] OR "unreliab*[tiab] OR "valid*[tiab] OR "coefficient of variation"[tiab] OR "coefficient"[tiab] OR "homogeneity"[tiab] OR "homogeneous"[tiab] OR "internal consistency"[tiab] OR ("cronbach*[tiab] AND ("alpha"[tiab] OR "alphas"[tiab])) OR ("item"[tiab] AND ("correlation*[tiab] OR "selection*[tiab] OR "reduction*[tiab])) OR "agreement"[tw] OR "precision"[tw] OR "imprecision"[tw] OR "precise values"[tw] OR "test-retest"[tiab] OR ("test"[tiab] AND "retest"[tiab]) OR ("reliab*[tiab] AND ("test"[tiab] OR "retest"[tiab])) OR "stability"[tiab] OR "interrater"[tiab] OR "inter-rater"[tiab] OR "intrarater"[tiab] OR "intra-rater"[tiab] OR "intertester"[tiab] OR "inter-tester"[tiab] OR "intratester"[tiab] OR "intra-tester"[tiab] OR "interobserver"[tiab] OR "inter-observer"[tiab] OR "intraobserver"[tiab] OR "intra-observer"[tiab] OR "intertechician"[tiab] OR "inter-technician"[tiab] OR "intratechnician"[tiab] OR "intra-technician"[tiab] OR "interexaminer"[tiab] OR "inter-examiner"[tiab] OR "intraexaminer"[tiab] OR "intra-examiner"[tiab] OR "interassay"[tiab] OR "inter-assay"[tiab] OR "intraassay"[tiab] OR "intra-assay"[tiab] OR "interindividual"[tiab] OR "inter-individual"[tiab] OR "intraindividual"[tiab] OR "intra-individual"[tiab] OR "interparticipant"[tiab] OR "inter-participant"[tiab] OR "intraparticipant"[tiab] OR "intra-participant"[tiab] OR "kappa"[tiab] OR "kappa's"[tiab] OR "kappas"[tiab] OR "repeatab*[tw] OR (("replicab*[tw] OR "repeated"[tw]) AND ("measure"[tw] OR "measures"[tw] OR "findings"[tw] OR "result"[tw] OR "results"[tw] OR "test"[tw] OR "tests"[tw])) OR "generaliza*[tiab] OR "generalisa*[tiab] OR "concordance"[tiab] OR ("intraclass"[tiab] AND "correlation*[tiab]) OR "discriminative"[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR "dimension*[tiab] OR "subscale*[tiab] OR ("multitrait"[tiab] AND "scaling"[tiab] AND ("analysis"[tiab] OR "analyses"[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*[tiab] OR "error"[tiab] OR "errors"[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR ("variability"[tiab] AND ("analysis"[tiab] OR "values"[tiab])) OR ("uncertainty"[tiab] AND ("measurement"[tiab] OR "measuring"[tiab])) OR "standard error of measurement"[tiab] OR "sensitiv*[tiab] OR "responsive*[tiab] OR ("limit"[tiab] AND "detection"[tiab]) OR "minimal detectable concentration"[tiab] OR "interpretab*[tiab] OR (("minimal"[tiab] OR "minimally"[tiab] OR "clinical"[tiab] OR "clinically"[tiab]) AND ("important"[tiab] OR "significant"[tiab] OR "detectable"[tiab]) AND ("change"[tiab] OR "difference"[tiab])) OR ("small*[tiab] AND ("real"[tiab] OR "detectable"[tiab]) AND ("change"[tiab] OR "difference"[tiab]))

OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR "IRT"[tiab] OR "Rasch"[tiab] OR "Differential item functioning"[tiab] OR "DIF"[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])

Block #3: Additional words specifically about diagnostic accuracy

OR ("index test"[tiab] OR "reference test"[tiab] OR "reference standard"[tiab] OR "gold standard"[tiab] OR "structured clinical interview"[tiab] OR "diagnostic accuracy"[tiab:~2] OR "DTA"[tiab] OR "predictive value"[tiab] OR "PPV"[tiab] OR "NPV"[tiab] OR "positive predictive value"[tiab] OR "negative predictive value"[tiab] OR "likelihood ratio"[tiab] OR "sensitivity"[ti] OR "specificity"[ti] OR "cut-off"[tiab] OR "cutoff"[tiab] OR "cut point"[tiab] OR "threshold score"[tiab] OR "ROC"[tiab] OR "receiver* operating characteristic"[tiab] OR "AUC"[tiab] OR "area* under the curve"[tiab] OR "case detection"[tiab] OR "case finding"[tiab] OR "classification"[tiab])

Block #4: Additional words about all measurement properties

OR ("Patient* experience"[tiab] OR "Patient* involvement"[tiab] OR "Patient* input"[tiab] OR "Patient* perception"[tiab] OR "User* feedback"[tiab] OR "Patient* feedback"[tiab] OR "Cognitive interview"[tiab] OR "Clinician* perception"[tiab] OR "Pilot testing"[tiab] OR "Field testing"[tiab] OR "Focus Groups"[Mesh] OR "Focus group"[tiab] OR "Cognitive debriefing"[tiab] OR "Expert* panel"[tiab] OR "Expert* review"[tiab] OR "Comprehensibility"[tiab] OR "Comprehensiveness"[tiab] OR "Item wording"[tiab] OR "Item evaluation"[tiab] OR "Item relevance"[tiab] OR "Item clarity"[tiab] OR "Item analy"[tiab] OR "Gold standard"[tiab] OR "Area under curve"[tiab] OR "Area under a curve"[tiab] OR "Translat*"[tiab] OR "Measurement invariance"[tiab] OR "Cross-cultural"[tiab] OR "Back-translation"[tiab] OR "Forward-translation"[tiab] OR "Cross-group equivalence"[tiab] OR "Linguistic validation"[tiab] OR "Measurement equivalence"[tiab] OR "Semantic equivalence"[tiab] OR "Conceptual equivalence"[tiab] OR "measure* propert*"[tiab] OR "CAT"[tiab] OR "computer adaptive testing"[tiab] OR "Scale development"[tiab] OR "Reflective model"[tiab] OR "Formative model"[tiab] OR "Measurement model"[tiab] OR "Classical test theory"[tiab] OR "Scoring algorithm"[tiab] OR "Instrument development"[tiab] OR "Questionnaire development"[tiab] OR "McDonald* omega"[tiab] OR "Omega coefficient"[tiab] OR "Item-total correlation"[tiab] OR "Reference score"[tiab] OR "Perceived change"[tiab] OR "smallest worthwhile effect"[tiab] OR "LoA"[tiab] OR "SDC"[tiab] OR "MIC"[tiab] OR "intraclass correlation coefficient"[tiab] OR "ICC"[tiab] OR "item characteristic curve"[tiab] OR (("eigenvalue"[tiab] OR "CFI"[tiab] OR "CFA"[tiab] OR "comparative fit index"[tiab] OR "TFI"[tiab] OR "tucker lewis index"[tiab] OR "RMSEA"[tiab] OR "root mean square error of approximation"[tiab] OR "SRMR"[tiab] OR "standardized root mean square residual"[tiab] OR "WRMR"[tiab] OR "weighted root mean square residual"[tiab] OR "ECV"[tiab] OR "explained common variance"[tiab] OR "PCA"[tiab] OR "component analy*"[tiab]) AND ("factor"[tiab] OR "item"[tiab])) OR "items factors"[tiab] OR "Item response theory"[tiab] OR "Kaiser criterion"[tiab] OR "Factor loadings"[tiab] OR "One-factor model"[tiab] OR "Two-factor model"[tiab] OR "Bifactor model"[tiab] OR "Bi-factor model"[tiab] OR "Localindependence"[tiab] OR "Monotonicity"[tiab] OR "Unidimensional*"[tiab] OR ("item"[tiab] AND ("correlation"[tiab] OR "selection"[tiab] OR "reduction"[tiab])) OR "Floor effect"[tiab] OR "Ceiling effect"[tiab] OR "hypothes* test"[tiab] OR "convergent"[tiab] OR "discriminant"[tiab] OR "known-groups"[tiab])

Block #5: COSMIN exclusion filter, with adaptations

NOT ("address"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lecture"[Publication Type] OR "legal case"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular work"[Publication Type] OR "congress"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type])

NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])

Block #6: Time limits

AND ("1995/01/01"[Date - Publication]: "2026/01/20"[Date - Publication])

Reference for the COSMIN filter:

Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, 18(8), 1115-1123.

Appendix D: Codebook for data extraction in Phase A (Living Mapping of the Evidence)

Category	Variable name	Variable definition	Permissible values
Study metadata	DOI	Digital Object Identifier	Alphanumeric DOI string
	Database identifier	Unique identifier assigned to the study by the database in which it is indexed (e.g., PubMed, Scopus, Web of Science, Embase)	PMID (numeric), Web of Science Accession Number (alphanumeric), Embase ID (alphanumeric), or another database-specific identifier.
	Author	Surname of the first author of the study	First author surname followed by <i>et al.</i> (if applicable)
	Year of publication	Calendar year in which the study was published	Four-digit year (e.g., 2018)
	Journal	Name of the journal in which the article was published	Alphanumeric
	Article ID	Unique identifier assigned to the article within the review	Review-assigned article ID (Alphanumeric string)
	Sample ID	Unique identifier assigned within the review to each unique sample in the article (if multiple samples)	Review-assigned sample ID Not applicable
Study characteristics	Study methodological approach	Study methodological approach based on the nature of the data collected and the analytic methods used.	Quantitative
			Qualitative
			Mixed-methods
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Study design	Type of methodological design used in the study	Observational – Cross-sectional
			Observational – Longitudinal Prospective
			Observational – Longitudinal Retrospective
			Observational – Longitudinal (Timing Not Reported)
			Interventional – Randomized Controlled Trial (RCT)

			Interventional – Non-randomized Trial /quasi-experimental
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Inclusion criteria with threshold	Indicate if the participants of the study were included according to a continuous threshold	Yes
			No
			Unclear – lack of reporting
			Unclear – don't know/don't understand
PROM characteristics	Patient-reported outcome measure (PROM) used	Name of the PROM used in the study that is within the scope of this review (If a study uses more than one eligible PROM, a separate data-extraction row will be completed for each PROM)	GAD7 (7-item anxiety scale)
			PHQ9 (9-item depression scale)
			RCADS 25
			RCADS 47
			WHODAS 36-item
			WHODAS 12-item (short form)
			WHODAS 12+24-item
	Scoring computation method (only for WHODAS)	Method used to calculate WHODAS 2.0 scores: simple (unweighted sum), complex (IRT-weighted)	Simple scoring
			Complex scoring
			Not applicable
			Unclear – lack of reporting
	Item wording modifications	Indicate whether the item wording has been modified from the original validated version of the instrument	Unclear – don't know/don't understand
			Yes
	Detail item wording modifications	Specify which changes were made to the item wording	No
			Alphanumeric string
	Response scoring modifications	Indicate whether the response scoring has been modified from the original validated version of the instrument	Not applicable
			Yes
	Detail response scoring modifications	Specify which changes were made to the response scoring	No
			Alphanumeric string
	Other PROM 1	Name of other instrument used in the study (to be completed if the study assessed any measurement property of another listed	Not applicable
			GAD2 (2-item anxiety screener)
			PHQ2 (2-item depression screener)
			PHQ4 (4-item combined anxiety/depression screener)

		instrument in addition to the main eligible PROM)	PHQ5 (5-item short form)
			PHQ8 (8-item depression scale (no suicidality item))
			RCADS 11
			RCADS 25
			RCADS 47
			WHODAS 36-item
			WHODAS 12-item (short form)
			WHODAS 12+24-item
			Other (specify)
			Not applicable
	Other PROM 1 (other)	Specify the name of the instrument when "Other (specify)" is selected for the variable Other PROM 1	Alphanumeric string
			Not applicable
	Other PROM 2	Name of other instrument used in the study (to be completed if the study assessed any measurement property of another listed instrument in addition to the main eligible PROM)	GAD2 (2-item anxiety screener)
			PHQ2 (2-item depression screener)
			PHQ4 (4-item combined anxiety/depression screener)
			PHQ5 (5-item short form)
			PHQ8 (8-item depression scale (no suicidality item))
			RCADS 11
			RCADS 25
			RCADS 47
			WHODAS 36-item
			WHODAS 12-item (short form)
			WHODAS 12+24-item
			Other (specify)
			Not applicable
	Other PROM 2 (other)	Specify the name of the instrument when "Other (specify)" is selected for the variable Other PROM 2	Alphanumeric string
			Not applicable
	Administration mode	Mode of administration indicating the informant of the scale	Self-reported version
			Caregiver report version (RCADS only)

			Interviewer version eliciting patient perspective
	Completion assistance	If the respondent requires assistance to complete the scale (without judgement or influence of the person assisting the application of the instrument), which type of assistance is provided to the target individual (informant) during completion of the scale	No assistance
			Interviewer-assisted
			Caregiver-assisted
			Clinician-assisted
			Other (specify)
	Completion assistance (other)	Description of the type of completion assistance when “Other (specify)” is selected for the variable Completion assistance	Alphanumeric string
			Not applicable
	Medium of administration	Format or medium used to deliver the scale	Paper and pencil
			Digital device (any)
			Telephone interview
			Other (specify)
			Unclear – lack of reporting
			Unclear – don’t know/don’t understand
	Medium of administration (other)	Description of the medium of administration when “Other (specify)” is selected for the variable Medium of administration	Alphanumeric string
			Not applicable
	Purpose of using the scale	Context of use of the PROM in which the measurement properties were evaluated	Diagnostic/screening
			Outcome assessment
			Other (specify)
			Unclear – lack of reporting
			Unclear – don’t know/don’t understand
	Language	Language in which the PROM is administered (as reported in the article)	Alphanumeric string
			Unclear – lack of reporting
			Unclear – don’t know/don’t understand
Participants’ characteristics	Description of the study sample	Description of the included study sample as shown in the article	Alphanumeric string
	Age group of the study sample	Age category of the study sample as reported in the article, based on the age range or mean age of participants	Adults
			Adolescents
			Adults and adolescents
			Unclear – lack of reporting

			Unclear – don't know/don't understand
Sample size	Total number of participants included in the study/analysis		Integer value (N)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Minimum age (inclusion criteria)	Lower eligible age threshold		Age in years
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Maximum age (inclusion criteria)	Upper eligible age threshold		Age in years
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Minimum age (study sample)	Youngest age among participants included in the study		Age in years
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Maximum age (study sample)	Oldest age among participants included in the study		Age in years
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Mean age of the study sample	Mean (average) age of the study sample		Numeric value (years)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Age standard deviation of the study sample	Standard deviation of age within the study sample		Numeric value (years)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Median age of the study sample	Median age within the study sample		Numeric value (years)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Number of male participants	Total number of male participants included in the study		Integer value (N)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Number of female participants	Total number of female participants included in the study		Integer value (N)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
Sex distribution	Percentage of study participants who are female, as reported or calculable from the study (use the proportion reported by the		Alphanumeric string. Percentage of male and female participants
			Unclear – lack of reporting

		authors or calculate it from provided sex distribution data when available)	Unclear – don't know/don't understand
	Gender distribution	Percentage of participants categorized by gender identity (e.g., man, woman, non-binary), if defined or implied in the study, based on self-identification or explicit gender measures	Alphanumeric string. Percentage of participants self-identified as man, woman, non-binary or other (specify the percentage for each reported gender).
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Type of population	General classification of the population studied	Clinical (physical health conditions)
			Clinical (mental health conditions)
			Clinical (substance-use disorders)
			Healthy volunteer (with no diagnosed conditions)
			General population (community sample)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Category physical health condition 1	Primary category describing the physical health condition present in the study population, if applicable	Cardiovascular diseases (e.g., hypertension, coronary artery disease, heart failure, stroke)
			Metabolic and endocrine disorders (e.g., diabetes, obesity, metabolic syndrome, thyroid disorders)
			Respiratory diseases (e.g., asthma, COPD)
			Neurological disorders (non-psychiatric) (e.g., Parkinson's disease, multiple sclerosis, epilepsy, migraine)
			Musculoskeletal conditions (e.g., arthritis, chronic back pain, osteoporosis)
			Chronic pain conditions (e.g., fibromyalgia, chronic low back pain, neuropathic pain)
			Cancer / oncological conditions
			Autoimmune and inflammatory diseases (e.g., rheumatoid arthritis, lupus, inflammatory bowel disease)

			Gastrointestinal diseases (e.g., IBS, Crohn's disease, ulcerative colitis)
			Renal and urological diseases
			Chronic infectious diseases (e.g., HIV, hepatitis)
			Sensory impairments (e.g., visual or hearing impairment)
			Other physical health condition
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Name of the physical health condition 1	Specific physical health diagnosis or condition present in the study population, as reported by the authors	Alphanumeric string (e.g., Type 2 Diabetes Mellitus)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Diagnostic basis physical health condition 1	Method used to determine the presence of the physical health condition among participants	Clinical diagnosis (medical record / clinician diagnosis)
			Standardized diagnostic criteria (e.g., ICD, ADA guidelines)
			Self-reported diagnosis
			Screening tool
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Definition physical health condition 1	Operational or clinical definition of the physical health condition used in the study, including diagnostic criteria, assessment tools, thresholds, or author-provided descriptions	Alphanumeric string
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Category physical health condition 2	Primary category describing an additional physical health condition present in the study population when multiple physical conditions	Cardiovascular diseases (e.g., hypertension, coronary artery disease, heart failure, stroke)

		are reported as comorbidities. If measurement properties are evaluated separately for different diseases using distinct participant samples, a separate data-extraction row will be completed for each health condition.	Metabolic and endocrine disorders (e.g., diabetes, obesity, metabolic syndrome, thyroid disorders)
			Respiratory diseases (e.g., asthma, COPD)
			Neurological disorders (non-psychiatric) (e.g., Parkinson's disease, multiple sclerosis, epilepsy, migraine)
			Musculoskeletal conditions (e.g., arthritis, chronic back pain, osteoporosis)
			Chronic pain conditions (e.g., fibromyalgia, chronic low back pain, neuropathic pain)
			Cancer / oncological conditions
			Autoimmune and inflammatory diseases (e.g., rheumatoid arthritis, lupus, inflammatory bowel disease)
			Gastrointestinal diseases (e.g., IBS, Crohn's disease, ulcerative colitis)
			Renal and urological diseases
			Chronic infectious diseases (e.g., HIV, hepatitis)
			Sensory impairments (e.g., visual or hearing impairment)
			Other physical health condition
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Name of the physical health condition 2	Specific physical health diagnosis or condition present in the study population, as reported by the authors	Alphanumeric string (e.g., Type 2 Diabetes Mellitus)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Diagnostic basis physical health condition 2		Clinical diagnosis (medical record / clinician diagnosis)

		Method used to determine the presence of the physical health condition among participants	Standardized diagnostic criteria (e.g., ICD, ADA guidelines)
			Self-reported diagnosis
			Screening tool
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Definition physical health condition 2	Operational or clinical definition of the physical health condition used in the study, including diagnostic criteria, assessment tools, thresholds, or author-provided descriptions	Alphanumeric string
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Category physical health condition 3	Primary category describing an additional physical health condition present in the study population when multiple physical conditions are reported as comorbidities. If measurement properties are evaluated separately for different diseases using distinct participant samples, a separate data-extraction row will be completed for each health condition	Cardiovascular diseases (e.g., hypertension, coronary artery disease, heart failure, stroke)
			Metabolic and endocrine disorders (e.g., diabetes, obesity, metabolic syndrome, thyroid disorders)
			Respiratory diseases (e.g., asthma, COPD)
			Neurological disorders (non-psychiatric) (e.g., Parkinson's disease, multiple sclerosis, epilepsy, migraine)
			Musculoskeletal conditions (e.g., arthritis, chronic back pain, osteoporosis)
			Chronic pain conditions (e.g., fibromyalgia, chronic low back pain, neuropathic pain)
			Cancer / oncological conditions
			Autoimmune and inflammatory diseases (e.g., rheumatoid arthritis, lupus, inflammatory bowel disease)
			Gastrointestinal diseases (e.g., IBS, Crohn's disease, ulcerative colitis)
			Renal and urological diseases

			Chronic infectious diseases (e.g., HIV, hepatitis)
			Sensory impairments (e.g., visual or hearing impairment)
			Other physical health condition
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Name of the physical health condition 3	Specific physical health diagnosis or condition present in the study population, as reported by the authors	Alphanumeric string (e.g., Type 2 Diabetes Mellitus)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Diagnostic basis physical health condition 3	Method used to determine the presence of the physical health condition among participants	Clinical diagnosis (medical record / clinician diagnosis)
			Standardized diagnostic criteria (e.g., ICD, ADA guidelines)
			Self-reported diagnosis
			Screening tool
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Definition physical health condition 3	Operational or clinical definition of the physical health condition used in the study, including diagnostic criteria, assessment tools, thresholds, or author-provided descriptions	Not applicable
			Alphanumeric string
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Category mental health condition 1	Primary category describing the mental health condition present in the study population, if applicable	Not applicable
			Depressive disorders (e.g., major depressive disorder, dysthymia)
			Anxiety disorders (e.g., generalized anxiety disorder, panic disorder, phobias)
			Stress-related and trauma-related disorders (e.g., PTSD, adjustment disorder)
			Bipolar and related disorders

			Psychotic disorders (e.g., schizophrenia, schizoaffective disorder)
			Neurodevelopmental disorders (e.g., ADHD, autism spectrum disorder)
			Eating disorders (e.g., anorexia nervosa, bulimia nervosa, binge-eating disorder)
			Personality disorders
			Cognitive disorders / dementia
			Sleep–wake disorders (e.g., insomnia)
			Other mental health condition
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Name of the mental health condition 1	Specific mental health diagnosis or condition present in the study population, as reported by the authors	Alphanumeric string (e.g., major depressive disorder)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Diagnostic basis mental health condition 1	Method used to determine the presence of the mental health condition among participants	Clinical diagnosis (medical record / clinician diagnosis)
			Standardized diagnostic criteria (e.g., DSM, ICD)
			Self-reported diagnosis
			Screening tool
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Definition mental health condition 1	Operational or clinical definition of the mental health condition used in the study, including diagnostic criteria, assessment tools, thresholds, or author-provided descriptions	Alphanumeric string
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Category mental health condition 2	Primary category describing an additional mental health condition present in the study	Depressive disorders (e.g., major depressive disorder, dysthymia)

		population when multiple mental health conditions are reported as comorbidities. If measurement properties are evaluated separately for different diseases using distinct participant samples, a separate data-extraction row will be completed for each health condition	Anxiety disorders (e.g., generalized anxiety disorder, panic disorder, phobias)
			Stress-related and trauma-related disorders (e.g., PTSD, adjustment disorder)
			Bipolar and related disorders
			Psychotic disorders (e.g., schizophrenia, schizoaffective disorder)
			Neurodevelopmental disorders (e.g., ADHD, autism spectrum disorder)
			Eating disorders (e.g., anorexia nervosa, bulimia nervosa, binge-eating disorder)
			Personality disorders
			Cognitive disorders / dementia
			Sleep–wake disorders (e.g., insomnia)
			Other mental health condition
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Name of the mental health condition 2	Specific mental health diagnosis or condition present in the study population, as reported by the authors	Alphanumeric string (e.g., major depressive disorder)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Diagnostic basis mental health condition 2	Method used to determine the presence of the mental health condition among participants	Clinical diagnosis (medical record / clinician diagnosis)
			Standardized diagnostic criteria (e.g., DSM, ICD)
			Self-reported diagnosis
			Screening tool
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Definition mental health condition 2	Operational or clinical definition of the mental health condition used in the study, including	Alphanumeric string
			Unclear – lack of reporting

		diagnostic criteria, assessment tools, thresholds, or author-provided descriptions	Unclear – don't know/don't understand
			Not applicable
	Category mental health condition 3	Primary category describing an additional mental health condition present in the study population when multiple mental health conditions are reported as comorbidities. If measurement properties are evaluated separately for different diseases using distinct participant samples, a separate data-extraction row will be completed for each health condition	Depressive disorders (e.g., major depressive disorder, dysthymia)
			Anxiety disorders (e.g., generalized anxiety disorder, panic disorder, phobias)
			Stress-related and trauma-related disorders (e.g., PTSD, adjustment disorder)
			Bipolar and related disorders
			Psychotic disorders (e.g., schizophrenia, schizoaffective disorder)
			Neurodevelopmental disorders (e.g., ADHD, autism spectrum disorder)
			Eating disorders (e.g., anorexia nervosa, bulimia nervosa, binge-eating disorder)
			Personality disorders
			Cognitive disorders / dementia
			Sleep-wake disorders (e.g., insomnia)
			Other mental health condition
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Name of the mental health condition 3	Specific mental health diagnosis or condition present in the study population, as reported by the authors	Alphanumeric string (e.g., major depressive disorder)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Diagnostic basis mental health condition 3	Method used to determine the presence of the mental health condition among participants	Clinical diagnosis (medical record / clinician diagnosis)
			Standardized diagnostic criteria (e.g., DSM, ICD)
			Self-reported diagnosis
			Screening tool
			Unclear – lack of reporting

			Unclear – don't know/don't understand
			Not applicable
	Definition mental health condition 3	Operational or clinical definition of the mental health condition used in the study, including diagnostic criteria, assessment tools, thresholds, or author-provided descriptions	Alphanumeric string
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Category substance-use disorder 1	Primary category describing the substance-use disorder present in the study population, if applicable	Alcohol use disorder
			Tobacco / nicotine use disorder
			Cannabis use disorder
			Opioid use disorder
			Stimulant use disorder (e.g., cocaine, amphetamines, methamphetamine)
			Sedative, hypnotic, or anxiolytic use disorder
			Polysubstance use disorder
			Other substance-use disorder
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Name of the substance-use disorder 1	Specific substance-use disorder present in the study population, as reported by the authors	Alphanumeric string (e.g., alcohol abuse disorder)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Diagnostic basis substance-use disorder 1	Method used to determine the presence of the substance-use disorder among participants	Clinical diagnosis (medical record / clinician diagnosis)
			Standardized diagnostic criteria (e.g., DSM, ICD)
			Self-reported diagnosis
			Screening tool
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
			Alphanumeric string

	Definition substance-use disorder 1	Operational or clinical definition of the substance-use disorder used in the study, including diagnostic criteria, assessment tools, thresholds, or author-provided descriptions	Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Category substance-use disorder 2	Primary category describing an additional substance-use disorder present in the study population when multiple substance-use disorders are reported as coexistent. If measurement properties are evaluated separately for different substance-use disorders using distinct participant samples, a separate data-extraction row will be completed for each disorder	Alcohol use disorder
			Tobacco / nicotine use disorder
			Cannabis use disorder
			Opioid use disorder
			Stimulant use disorder (e.g., cocaine, amphetamines, methamphetamine)
			Sedative, hypnotic, or anxiolytic use disorder
			Polysubstance use disorder
			Other substance-use disorder
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Name of the substance-use disorder 2	Specific substance-use disorder present in the study population, as reported by the authors	Alphanumeric string (e.g., alcohol abuse disorder)
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Diagnostic basis substance-use disorder 2	Method used to determine the presence of the substance-use disorder among participants	Clinical diagnosis (medical record / clinician diagnosis)
			Standardized diagnostic criteria (e.g., DSM, ICD)
			Self-reported diagnosis
			Screening tool
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Not applicable
	Definition substance-use disorder 2	Operational or clinical definition of the substance-use disorder used in the study, including diagnostic criteria, assessment	Alphanumeric string
			Unclear – lack of reporting
			Unclear – don't know/don't understand

		tools, thresholds, or author-provided descriptions	Not applicable
Category substance-use disorder 3	Primary category describing an additional substance-use disorder present in the study population when multiple substance-use disorders are reported as coexistent. If measurement properties are evaluated separately for different substance-use disorders using distinct participant samples, a separate data-extraction row will be completed for each disorder	Alcohol use disorder	
		Tobacco / nicotine use disorder	
		Cannabis use disorder	
		Opioid use disorder	
		Stimulant use disorder (e.g., cocaine, amphetamines, methamphetamine)	
		Sedative, hypnotic, or anxiolytic use disorder	
		Polysubstance use disorder	
		Other substance-use disorder	
		Unclear – lack of reporting	
		Unclear – don't know/don't understand	
		Not applicable	
Name of the substance-use disorder 3	Specific substance-use disorder present in the study population, as reported by the authors	Alphanumeric string (e.g., alcohol abuse disorder)	
		Unclear – lack of reporting	
		Unclear – don't know/don't understand	
		Not applicable	
Diagnostic basis substance-use disorder 3	Method used to determine the presence of the substance-use disorder among participants	Clinical diagnosis (medical record / clinician diagnosis)	
		Standardized diagnostic criteria (e.g., DSM, ICD)	
		Self-reported diagnosis	
		Screening tool	
		Unclear – lack of reporting	
		Unclear – don't know/don't understand	
		Not applicable	
Definition substance-use disorder 3	Operational or clinical definition of the substance-use disorder used in the study, including diagnostic criteria, assessment tools, thresholds, or author-provided descriptions	Alphanumeric string	
		Unclear – lack of reporting	
		Unclear – don't know/don't understand	
		Not applicable	
Specific demographics		Veterans	

		Identification of specific demographic subgroups targeted in the study	LGBTQIA+
			Migrants
			Other (specify)
	Geographical setting	Geographic location where the study was conducted	Country or countries
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Healthcare setting level	Level of healthcare system in which the study was conducted	Primary care
			Secondary care
			Tertiary care
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Patient setting	Care context in which participants received services or were studied	Not applicable
			Outpatient
			Inpatient
			Unclear – lack of reporting
			Unclear – don't know/don't understand
PROM construct and assessed measurement properties in the study	PROM construct	Construct that the PROM is intended to assess, as explicitly stated by the study authors	Alphanumeric string. Construct description using authors' wording (e.g., quality of life, physical functioning, pain intensity, mental health, symptom severity, etc.)
	Content validity*	The degree to which the content of a PROM is an adequate reflection of the construct to be measured.	Yes – reported results
			Not assessed
			Unclear – lack of reporting
	Structural validity*	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured.	Unclear – don't know/don't understand
			Yes – reported results
			Not assessed
	Internal consistency*	The degree of the interrelatedness among the items.	Unclear – lack of reporting
			Unclear – don't know/don't understand
			Yes – reported results
			Yes – reported results

	Cross-cultural validity / measurement invariance*	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM.	Not assessed
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Reliability*	The extent to which scores for respondents who have not changed are the same for repeated measurement under several conditions: e.g., using different sets of items from the same PROM (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e., raters or responders) on different occasions (intra-rater).	Yes – reported results
			Not assessed
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Measurement error*	The systematic and random error of a respondent's score that is not attributed to true changes in the construct to be measured.	Yes – reported results
			Not assessed
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Criterion validity*	Criterion validity refers to the degree to which PROM scores are an adequate reflection of a reference standard (gold standard).	Yes – reported results
			Not assessed
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Comparator instrument for criterion validity	If results for criterion validity are reported, specify which instrument has been used as comparator	Alphanumeric string.
	Hypothesis testing for construct validity*	The degree to which the scores of a PROM are consistent with hypotheses (with regard to relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the PROM validly measures the construct to be measured.	Not applicable
			Yes – reported results
			Not assessed
			Unclear – lack of reporting
			Unclear – don't know/don't understand
			Alphanumeric string.

	Comparator instrument for hypothesis testing for construct validity	If results for hypothesis testing for construct validity are reported, specify which instrument has been used as comparator	Not applicable
	Responsiveness*	The ability of a PROM to detect change over time in the construct to be measured.	Yes – reported results
			Not assessed
			Unclear – lack of reporting
			Unclear – don't know/don't understand
	Comparator instrument for Responsiveness	If results for Responsiveness are reported, specify which instrument has been used as comparator	Alphanumeric string.
Other properties of the measurement tools	Feasibility*	Ease of application of the PROM in its intended context of use, given constraints such as time or money.	Not applicable
			Yes – reported results
			Not assessed
			Unclear – lack of reporting
	Interpretability*	The degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations – to a PROM's quantitative scores or change in scores.	Unclear – don't know/don't understand
			Yes – reported results
			Not assessed
			Unclear – lack of reporting
	Acceptability	How respondents perceive the experience of completing the scale, e.g., perceived intrusiveness, burden, potential harms or benefits.	Unclear – don't know/don't understand
			Yes – reported results
			Not assessed
			Unclear – lack of reporting
Involvement of contributors with lived experiences	Involvement of lived-experience contributors	Indicate whether individuals with lived experience actively contributed to the study.	Unclear – don't know/don't understand
			Yes
			No
			Unclear – lack of reporting

* Definition derived from the COSMIN taxonomy, as specified in the COSMIN 2.0 guidelines (Mokkink et al., 2024).