INPLASY

INPLASY202570014

doi: 10.37766/inplasy2025.7.0014

Received: 3 July 2025

Published: 3 July 2025

Corresponding author:
Klaus Linde

klaus.linde@mri.tum.de

Author Affiliation:
Institute of General Practice and Health Services Research, TUM School of Medicine and Health, Department of Clinical Medicine, Technical University of Munich, Munich, Germany.

Comparative test accuracy of eight self-rating scales for screening for anxiety disorders among adults: protocol for a network meta-analysis

Linde. K; Cerullo, E; Fomenko, A; Aktürk, Z; Eck, S; Dümmler, D; Jones, HE; Derezea, E; Hapfelmeier, A; Sutton, A; Schneider, A.

ADMINISTRATIVE INFORMATION

Support - There is no external funding for the planned project. However, part of the work on which the project described below is based (systematic reviews 1 to 4 - see section 'Review question/objective') was funded by the German Federal Ministry of Education, Research Grant 01KG2105. There was no external funding for reviews 5 and 6.

Review Stage at time of this submission - Data analysis.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202570014

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 3 July 2025 and was last updated on 3 July 2025.

INTRODUCTION

Review question / Objective The objective of the planned network meta-analysis (NMA) is to estimate and compare the accuracy of each of eight self-report screening tests for anxiety disorders.

This is a protocol and analysis plan for a test accuracy NMA, based on a series of six systematic reviews of the accuracy of eight anxiety self-rating scales:
- Review 1: Generalized Anxiety Disorder 7-item Scale (GAD-7) and Generalized Anxiety Disorder 2-item Scale (GAD-2) [1]
- Review 2: Hospital Anxiety and Depression Scale – Anxiety subscale (HADS-A) [2]
- Review 3: Beck Anxiety Inventory (BAI) – under review [3]
- Review 4: State-Trait Anxiety Inventory – Trait (STAI-T) and State-Trait Anxiety Inventory – State (STAI-S) – under review [4]
- Review 5: Overall Anxiety Severity and Impairment Scale (OASIS) – submitted for publication [5]
- Review 6: Patient-Reported Outcomes Measurement Information System (PROMIS) Anxiety Short Forms – in manuscript preparation

The joint protocol for reviews 1 to 4 has been published previously [6], and the protocol for reviews 5 and 6 has been registered in PROSPERO [7]. For the planned NMA, the data collected in the six reviews have been merged into a joint database. No new data will be collected. No new selection criteria will be applied. The basic idea of the NMA was already described in the first protocol [6]. Here we add detail to the planned analyses before commencing the analyses.

References:
1. Aktürk Z et al. Generalized Anxiety Disorder 7-item (GAD-7) and 2-item (GAD-2) scales for detecting anxiety disorders in adults. Cochrane

Database of Systematic Reviews. 2025 Mar 25;3(3):CD015455.
2. Fomenko A, et al. Hospital Anxiety and Depression Scale Anxiety subscale. Cochrane Database of Systematic Reviews 2025, Issue 7. Art. No.: CD015456.
3. Eck S et al. BeckAnxiety Inventory (BAI) for detecting anxiety disorders in adults. Cochrane Database of Systematic Reviews (in peer review).
4. Dümmler D et al. State-Trait Anxiety Inventory (STAI) for detecting anxiety disorders in adults. Cochrane Database of Systematic Reviews (in peer review).
5. Fomenko A et al. Test accuracy of the Overall Anxiety Severity and Impairment Scale (OASIS) for screening anxiety in adults – systematic review and multiple thresholds meta-analysis (submitted for publication).
6. Linde K et al. The diagnostic accuracy of widely used self-report questionnaires for detecting anxietydisorders in adults. Cochrane Database of Systematic Reviews 2022, Issue 9. Art. No: CD015292.
7. Fomenko A et al. The diagnostic accuracy of self-report questionnaires OASIS and PROMIS Anxiety Short Form for detecting anxiety disorders in adults. PROSPERO 2023 CRD42023485827. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023485827.

Rationale Anxiety disorders are frequent but often remain undetected, even in persons in which treatment would be necessary. Some expert panels have recommended screening for anxiety disorders with self-report questionnaires (e.g. [8]). A number of such questionnaires have been investigated in test accuracy studies. In a series of reviews, we have summarized the available studies for eight questionnaires and estimated their sensitivity and specificity when used as screeners. However, since there are no head-to-head comparisons of all questionnaires in the existing literature, these reviews cannot provide a clear answer on the question whether some questionnaires are more accurate than others. In this situation, a NMA is a logical next step [9].

References
8. US Preventive Services Task Force. Screening for Anxiety Disorders in Adults: US Preventive Services Task Force Recommendation Statement. JAMA 2023;329(24):2163-2170.
9. Veroniki AA et al. Diagnostic test accuracy network meta-analysis methods: A scoping review and empirical assessment. J Clin Epidemiol. 2022;146:86-96.

Condition being studied Anxiety disorders are classified in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV or DSM-5) [10,11] and the International Statistical Classification of Diseases and Related Health Conditions (ICD-10 or ICD-11) [12,13]. Important entities of anxiety disorders are generalised anxiety disorder (GAD), agoraphobia, specific phobia, social anxiety disorder (social phobia), and separation anxiety. Obsessive-compulsive disorder and post-traumatic stress disorder had been included in former classifications but were removed from anxiety disorders with the introduction of DSM-5. Studies using classifications before DSM IV or ICD-10 were not included because older diagnostic criteria were different.

The primary target condition for our NMA will be 'any anxiety disorder'. This summary category is particularly relevant for screening purposes as the prevalence of specific anxiety disorders is usually very low, making screening with self-report questionnaires less efficient. Yet, we will also investigate 'generalized anxiety disorder' as a secondary target condition as it is the most widely studied specific anxiety disorder in the available studies.

References
10. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). Washington, DC: American Psychiatric Association, 2000.
11. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).Arlington, VA: American Psychiatric Association, 2013.
12. World Health Organization. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. 2 edition. World Health Organization, 1993.
13. World Health Organization. ICD-11 for Mortality and Morbidity Statistics: Mental, behavioural or neurodevelopmental disorders. World Health Organization, 2019.

METHODS

Search strategy No new searches will be done for this network meta-analysis. For the six reviews mentioned above, we searched the following databases using relevant subject headings (controlled vocabularies), text-words and search syntax appropriate to each resource:
- Embase (Ovid)
- MEDLINE (Ovid)
- PubMed-NOT-MEDLINE (NLM)
- PsycINFO (Ovid)

We did not apply any restrictions on language or publication status to the searches. The full overall search strategy for EMBASE has been reported in the first protocol [6]. The search strategies adapted for the individual questionnaires are reported in the respective reviews [1-5]. The latest update searches for each of the six reviews were performed in July 2024.

Participant or population We included studies in adults (mean age 18 years or older) screened for anxiety disorders. We excluded studies on patients seeking help in mental health settings or patients who were recruited specifically due to mental health symptoms in other settings as this does not correspond to a traditional screening approach.

Intervention To be included in the data set, a study must have evaluated the accuracy of at least one of the following eight self-assessment questionnaires (index tests): Generalized Anxiety Disorder 7-item Scale (GAD-7) [14], Generalized Anxiety Disorder 2-item Scale (GAD-2) [14], Hospital Anxiety and Depression Scale – Anxiety subscale (HADS-A) [15], Beck Anxiety Inventory (BAI) [16], State-Trait Anxiety Inventory – Trait (STAI-T) [17], State-Trait Anxiety Inventory – State (STAI-S) [17], Overall Anxiety Severity and Impairment Scale (OASIS) [18], or Patient-Reported Outcomes Measurement Information System (PROMIS) Anxiety Short Form 8a [19]. There are a number of other questionnaires which could not be investigated due to limited resources and the small number of existing test accuracy studies. Criteria for focussing on the eight questionnaires selected were number of existing test accuracy studies in adults (based on preliminary searches before starting the project), frequency of clinical use and availability in multiple languages.

References

14. Kroenke K et al. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Annals of Internal Medicine 2007;146(5):317-25.
15. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatrica Scandinavica 1983;67(6):361-70.
16. Beck AT et al. An inventory for measuring clinical anxiety: psychometric properties. Journal of Consulting and Clinical Psychology 1988;56(6):893-7.
17. Spielberger et al. STAI Manual for the State-Trait Anxiety Inventory. Palo Alto (CA): Consulting Psychologists' Press. Inc, 1970.
18. Norman SB et al. Development and validation of an Overall Anxiety Severity And Impairment Scale (OASIS). Depress Anxiety. 2006;23(4):245-9.
19. Pilkonis PA et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment. 2011 Sep;18(3):263-83.

Comparator The diagnosis of anxiety of all participants in primary studies must have been made or ruled out using a validated structured or semi-structured clinical interview as reference standard, such as the SCID (Structured Clinical Interview for DSM), the CIDI (Composite International Diagnostic Interview), the MINI (Mini-International Neuropsychiatric Interview), DIA-X (diagnostic expert system for mental disorders); SADS (Schedule for Affective Disorders and Schizophrenia) and DIPS, or Mini-DIPS (German: "Diagnostisches Interview bei psychischen Störungen"). We excluded studies in which the diagnosis was informally based on a checklist based on ICD or DSM or a clinical diagnosis without operationalisation or on another questionnaire as this is not considered an adequate and reliable reference standard for mental health studies.

Study designs to be included We included all studies that allowed for the construction of at least one 2 x 2 table based on index test and reference standard results (number of true positive, false positive, false negative and true negative index test results). We included cross-sectional studies and studies of longitudinal design, in which the index test and the reference standard were applied cross-sectionally. We excluded case-control studies in which cases suffered from anxiety.

Eligibility criteria The main eligibility criteria were described above. We did not apply any restrictions regarding language, publication type or sample size.

Information sources We searched Embase (Ovid), MEDLINE (Ovid), PubMed-NOT-MEDLINE (NLM), PsycINFO (Ovid). In addition, we screened references of included studies and relevant reviews for potentially relevant studies. For further details, see section protocol [6] and our published reviews [1,2].

Main outcome(s) The area under the summary Receiver Operating Characteristic (sROC) curve (AUC) will be used as a single number summary of the overall accuracy of each test. As well as providing 95% credible intervals (CrI) and 95%

prediction intervals (PrI) around each questionnaire's estimated AUC, the relative ranking of these performance measures will also be estimated (from highest to lowest) and a 95% CrI and 95% PrI constructed for each of these ranks to indicate where important differences do and do not exist. We will also report differences in AUC's for each pair of tests along with 95% CrI's of these differences.

Additional outcome(s) The fitted meta-analysis model will additionally enable estimation of summary test accuracy, in terms of sensitivity (Se) and specificity (Sp) together with 95% CrIs and 95% PrIs, to be calculated for each threshold of each scale included in the analysis.
The numerical threshold that each point on the sROC curves (along with the 95% CrIs and PrIs) corresponds to will be presented visually on sROC plots. This will allow the comparison of accuracy of any of the included tests at any of their thresholds.
The summary sensitivity and specificity for each test at particular thresholds will also be reported:
i) At the threshold which maximises the Youden index on each scale;
ii) At test threshold values which are most commonly used in the literature for each test.

Furthermore, we will make formal comparisons between the sensitivity and specificity of each test, at the most commonly used threshold for each test. These comparisons will be reported using (1) ranked estimates of sensitivity and specificity, across all tests, and (2) estimates of the difference in sensitivity and specificity (with 95% CrI) for each pair of tests. For tests with multiple commonly used thresholds, we will present results for each threshold to show how threshold choice affects test performance and comparative rankings.

Data management At least two reviewers independently performed the following steps: screening of and selection of studies (using EPPI Reviewer in reviews 1 to 4 and Rayyan for reviews 5 and 6); data extraction (using pre-tested Excel spreadsheets; for details, see [6]); and quality assessment. Study authors were systematically contacted for additional information relevant to the quality assessments and for additional data if test accuracy findings were not presented for all thresholds.
For the NMA, relevant variables of the reviews were merged into a common Excel file.

Quality assessment / Risk of bias analysis Risk of bias and the external validity of the included studies were assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [20]. Following the recommendations of the Cochrane Diagnostic Test Accuracy Working Group [21], we developed coding guidelines for each item (see [1,2,6] for details).

References
20. Whiting PF et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine 2011;155(8):529-36.
21. Davenport CF et al. Use of QUADAS-2. Lesson 6.3: Cochrane Collaboration DTA Online Learning Materials. The Cochrane Collaboration September 2012 (available at http://training.cochrane.org).

Strategy of data synthesis We will use the dplyr [22] R package to clean and standardize the extracted data from the consolidated Excel files (see Data Management section). The raw data contains study-level test accuracy estimates (Se, Sp, true positives, false positives, etc. (directly derived from the standard 2 x 2 test comparison tables)) reported at various thresholds for each anxiety screening instrument.
From this tidy dataset, we will construct cumulative count matrices for meta-analysis using a custom R function. For each anxiety screening test, this function creates two matrices (for the diseased and non-diseased populations), where each row represents a study and each column represents a threshold, from 0 to the maximum possible score for that instrument. Matrix entries contain the number of individuals scoring at or above each threshold: these are the numbers of true positives (TP) for the diseased population and false positives (FP) for the non-diseased population.
These cumulative counts are monotonically non-increasing across thresholds within each study, reflecting the constraint that fewer individuals test positive as thresholds become more stringent. Missing threshold data (where studies did not report results at specific cutoffs) are coded as -1, allowing inclusion of studies reporting at different or incomplete threshold sets into the statistical models.
For the data analysis, we will use an ordinal-bivariate network meta-analysis (NMA) diagnostic test accuracy (NMA-DTA) model recently proposed by Cerullo et al [23], which is a "multiple thresholds" model based on ordinal regression. It can be considered to be an ordinal extension of the standard bivariate model [24], and also draws upon some features of a multiple thresholds model for a single continuous index test proposed by Jones et al [25] and an NMA-DTA model proposed by Nyaga et al [26].

To fit these models, we will use the MetaOrdDTA R package [27], which uses the Bayesian MCMC software Stan (Stan development team, 2024 [28]). We will fit all models using the default "diffuse" prior distributions specified in the package.
We will fit the ordinal-bivariate NMA-DTA model from the MetaOrdDTA R package - since all of the index tests are ordinal (i.e., GAD-2 [6 thresholds], GAD-7 [21 thresholds], HADS [21 thresholds], OASIS [20 thresholds], and PROMIS-SF-v1.0-Anxiety-8a [32 thresholds], BAI [63 thresholds], STAI-S [60 thresholds], and STAI-T [60 thresholds]). Furthermore, we are choosing to use the ordinal-bivariate NMA-DTA model because Cerullo et al [27] recently conducted a comprehensive simulation study which showed that this model outperformed other competing models for this kind of ordinal test accuracy data.
We will fit this model using both fixed-effects and random-effects thresholds. The choice of whether results from the fixed-effect or random-effects threshold models are presented as the "final" results will be based on model fit (penalised for complexity). Similarly, the choice of whether we assume that the between-study variances are homogenous across tests (i.e. compound symmetry matrix structure) or heterogenous across tests (i.e. unstructured covariance matrix) will also be decided based on model fit.

References see section sensitivity analysis.

Subgroup analysis Rather than split the data and conduct separate analyses on each subgroup, we will explore the impact of binary and categorical study-level covariates by including (meta-) regression terms in the analysis models. This will allow the exploration of whether and how results differ by covariate and provide a quantitative estimate of any differences, with 95% CrIs. It will also allow the quantification of the reduction in heterogeneity associated with each covariate, via percentage reduction in the estimated standard deviations of random effects.
We will perform univariable meta-regression for the following covariates:
1. clinical setting [categorical covariate]
2. type of reference standard [categorical covariate]
3. logit- (true) prevalence of target condition [continuous covariate]
Furthermore, we will consider common, independent and exchangeable effects for the effect of each covariate on each test. Choice of which of these results to present will be informed by model fit statistics, penalised for complexity.
If the univariable meta-regression results suggest that multiple covariates seem important, we will fit a final multivariable model and present results for this as well.

Sensitivity analysis Sensitivity analysis will be used to assess the robustness of the results to analysis decisions. This will include considering the impact of study quality on the results by restricting analysis to only studies at low risk of bias, based on the QUADAS-2 assessment. If there is insufficient data (i.e., number of studies) to perform this, we will instead explore the impact of study quality using meta-regression, with high vs. low risk of bias as a covariate.
We will also investigate whether the exclusion of studies for which no test accuracy data was published but only obtained from authors (usually studies which did not originally aim to investigate test accuracy) has an impact on findings.

References
22. Wickham H et al (2023). dplyr: a grammar of data manipulation. R package version 1.1.4, https://github.com/tidyverse/dplyr, https://dplyr.tidyverse.org
23. Cerullo E et al. Ordinal regression for meta-analysis of test accuracy: a flexible approach for utilising all threshold data. 29 May 2025, arXiv:2505.23393v1
24. Reitsma JB et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005 Oct;58(10):982-90.
25. Jones HE et al. Quantifying how diagnostic test accuracy depends on threshold in a meta-analysis. Stat Med. 2019 Oct 30;38(24):4789-4803. doi: 10.1002/sim.8301. Epub 2019 Sep 30. Erratum in: Stat Med. 2021 Aug 15;40(18):4166.
26. Nyaga VN et al. ANOVA model for network meta-analysis of diagnostic test accuracy data. Stat Methods Med Res. 2018 Jun;27(6):1766-1784.
27. Cerullo E. (2025). MetaOrdDTA: An R package for the meta-analysis and network meta-analysis of medical tests across all thresholds. R package version 0.1.0. https://github.com/CerulloE1996/MetaOrdDTA
28. Stan Development Team. 2024. Stan Reference Manual, v2.36.0. https://mc-stan.org
29. Rutter et al. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001 Oct 15;20(19):2865-84.

Language restriction There are no language restrictions.

Country(ies) involved Germany, UK.

Keywords Diagnostic test accuracy, screening, anxiety disorders, self-rating questionnaires, network meta-analysis.

Dissemination plans The findings of the NMA will be submitted to peer reviewed international medical journals.

Contributions of each author

Author 1 - Klaus Linde - planning, study selection, data extraction, quality assessment, interpretation, overall coordination.
Email: klaus.linde@mri.tum.de
Author 2 - Enzo Cerullo - statistical analysis, interpretation.
Email: ec325@leicester.ac.uk
Author 3 - Alexey Fomenko - study selection, data extraction, quality assessment, interpretation.
Email: alexeyaleksandrovic.fomenko@mri.tum.de
Author 4 - Zekeriya Aktürk - study selection, data extraction, quality assessment, interpretation.
Email: zekeriya.aktuerk@med.uni-augsburg.de
Author 5 - Stefanie Eck - study selection, data extraction, quality assessment, interpretation.
Email: stefanie.eck@mri.tum.de
Author 6 - Daniel Dümmler - study selection, data extraction, quality assessment, interpretation statistical analysis, interpretation.
Email: daniel.duemmler@uni-muenster.de
Author 7 - Hayley E Jones - statistical analysis, interpretation.
Email: hayley.jones@bristol.ac.uk
Author 8 - Efthymia Derezea - statistical analysis, interpretation.
Email: e.derezea@bristol.ac.uk
Author 9 - Alexander Hapfelmeier - statistical analysis, interpretation.
Email: alexander.hapfelmeier@mri.tum.de
Author 10 - Alex J Sutton - statistical analysis, interpretation.
Email: ajs22@leicester.ac.uk
Author 11 - Antonius Schneider - study idea, clinical expertise, interpretation, overall supervision.
Email: antonius.schneider@mri.tum.de