Comparison of the accuracy of the 7-item HADS Depression subscale and 14-item total HADS for screening for major depression: a systematic review and individual participant data meta-analysis

Abstract

The 7-item Hospital Anxiety and Depression Scale Depression subscale (HADS-D) and the total score of the 14-item HADS (HADS-T) are both used for major depression screening. Compared to the HADS-D, the HADS-T includes anxiety items and requires more time to complete. We compared the screening accuracy of the HADS-D and HADS-T for major depression detection. We conducted an individual participant data meta-analysis and fit bivariate random-effects models to assess diagnostic accuracy among participants with both HADS-D and HADS-T scores. We identified optimal cutoffs, estimated sensitivity and specificity with 95% confidence intervals (CIs), and compared screening accuracy across paired cutoffs via two-stage and individual-level models. We used a 0.05 equivalence margin to assess equivalency in sensitivity and specificity. 20,700 participants (2,285 major depression cases) from 98 studies were included. Cutoffs of ≥ 7 for the HADS-D (sensitivity 0.79 [0.75, 0.83], specificity 0.78 [0.75, 0.80]) and ≥ 15 for the HADS-T (sensitivity 0.79 [0.76, 0.82], specificity 0.81 [0.78, 0.83]) minimized the distance to the top-left corner of the receiver operating characteristic curve. Across all sets of paired cutoffs evaluated, differences of sensitivity between HADS-T and HADS-D ranged from -0.05 to 0.01 (0.00 at paired optimal cutoffs), and differences of specificity were within 0.03 for all cutoffs (0.02 to 0.03). The pattern was similar among outpatients, although the HADS-T was slightly (not non-equivalently) more specific among inpatients. The accuracy of HADS-T was equivalent to the HADS-D for detecting major depression. In most settings, the shorter HADS-D would be preferred.

Keywords: HADS-D, HADS-T, individual participant data meta-analysis, depression screening, diagnostic accuracy

Public significance statements:

The present study suggests that the accuracy of 14-item Hospital Anxiety and Depression Scale (HADS-D) and the 7-item HADS Depression subscale (HADS-D) are equivalent for detecting major depression. Using the 7-item HADS-D for depression screening instead of the full 14-item HADS-T has minimal influence on performance of the measure but would reduce patient and participant burden in most clinical and research settings.

The 14-item Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983) was developed to facilitate the identification of anxiety disorders and major depression in people with a physical illness. The HADS includes two subscales. The 7-item Depression subscale (HADS-D) was designed to assess continuous depressive symptoms and for depression screening, whereas the 7-item Anxiety subscale (HADS-A) was designed to assess and screen for anxiety (Zigmond & Snaith, 1983). Both HADS-D and full HADS total scores (HADS-T) have been used to screen for major depression (Mitchell, Meader, & Symonds, 2010; Vodermaier & Millman, 2011). The HADS-T takes more time to complete and includes anxiety items not specific to depression. Some have suggested, though, that anxiety symptoms should be considered when assessing depression (Schatzberg, 2019). Furthermore, previous reviews have provided some preliminary evidence that HADS-T may perform better than the HADS-D (Mitchell, Meader, & Symonds, 2010; Vodermaier & Millman, 2011).

Commonly used HADS-D cutoff thresholds of ≥ 8 for "possible" depression and ≥ 11 for "probable" depression were established in the original validation study, which included only 100 participants (11 depression cases) (Zigmond & Snaith, 1983). A recent individual participant data meta-analysis (IPDMA) on HADS-D accuracy to screen for major depression (101 studies; 22,574 participants; 2,549 major depression cases) found that a cutoff of ≥ 7 maximized combined sensitivity and specificity across reference standards; standard cutoffs of ≥ 8 and ≥ 11 were less sensitive but more specific (Wu, Levis, Sun, et al., 2021). There is not a standard cutoff for screening to detect major depression with the HADS-T.

Two previous meta-analyses, both done with studies of cancer patients, have indirectly compared the HADS-D and HADS-T for detecting major depression (Mitchell et al., 2010; Vodermaier & Millman, 2011). Both searched through October 2009 for eligible studies. One evaluated 9 studies that used the HADS-D with a cutoff of 8 or greater and 6 studies that used the HADS-T with a cutoff of 15 (number of participants not reported) (Mitchell et al., 2010), whereas the other included 2-5 studies each in analyses of HADS-D cutoffs of 7, 9, and 11 and HADS-T cutoffs of 15, 17, 19 and 20 (470 to 872 participants per analysis) (Vodermaier & Millman, 2011). Both meta-analyses suggested that the HADS-T may perform better than the HADS-D, but there was a high level of uncertainty due to indirect comparisons between participants from different studies that reported HADS-D and HADS-T results, the small number of total participants, and possible selective outcome reporting bias (Levis et al., 2017; Neupane et al., 2021; Rice & Thombs, 2016; Thombs et al., 2011; Thombs & Rice, 2016) since not all primary studies reported results from the same cutoffs.

Using the full 14-item HADS-T for depression screening would be warranted if it is sufficiently more accurate than the shorter 7-item HADS-D to justify the additional time and patient burden involved. We previously assessed the accuracy of the HADS-D using IPDMA (Wu, Levis, Sun, et al., 2021). IPDMA involves a standard systematic review, followed by synthesis of original research data from primary studies, rather than extracting summary data (Riley, Lambert, & Abo-Zaid, 2010). In that IPDMA, we found that diagnostic accuracy of HADS-D was not significantly different for any cutoffs across reference standards based on participant characteristics, including age, sex, cancer diagnosis, country human development index levels, participant recruitment settings, or the study’s risk of bias ratings (Wu et al., 2021). In the present study, we included studies from the HADS-D IPDMA where HADS-T scores were provided or could be calculated from individual item scores. Our objectives were to (1) directly compare screening accuracy of the HADS-T and HADS-D for major depression detection using the same participant data across all studies regardless of reference standard, and (2) replicate the comparison among studies that used a semi-structured diagnostic interview [e.g., Structured Clinical Interview for the DSM (SCID) (First, 1995)] as a reference standard, since semi-structured interviews more closely reflect the actual diagnostic process than fully-structured interviews.

Methods

The present study used a subset of studies and participants from our previously conducted HADS-D IPDMA (Wu, Levis, Sun, et al., 2021) for which HADS-T scores were also available. Analyses of HADS-D and HADS-T diagnostic accuracy were conducted according to the HADS-D IPDMA methods (Wu, Levis, Sun, et al., 2021) with the addition of analyses to directly compare HADS-D and HADS-T accuracy.

Dataset eligibility

For the main HADS-D meta-analysis, datasets from articles in any language were eligible for inclusion if (1) they included diagnostic classification for current Major Depressive Disorder (MDD) or Major Depressive Episode (MDE) using Diagnostic and Statistical Manual of Mental Disorders (DSM) (American Psychiatric Association, 1987; 1994; 2000; 2013) or International Classification of Diseases (ICD) (World Health Organization (WHO), 1992) criteria based on a validated semi-structured or fully structured interview; (2) they included total scores for the HADS-D; (3) the diagnostic interview and HADS-D were administered within two weeks of each other, because DSM and ICD major depression diagnostic criteria specify that symptoms must have been present in the last two weeks; (4) participants were ≥ 18 years of age and not recruited from youth or psychiatric settings; and (5) participants were not recruited because they were identified as having symptoms of depression, since screening is done to identify previously unrecognized cases. We focused on MDD and MDE because major guidelines on depression screening have focused on screening for major depression but have not considered screening for less severe conditions, such as dysthymia or persistent depressive disorder, for which treatment options and effectiveness are much less well delineated (Joffres et al., 2013; National Collaborating Centre for Mental Health (UK), 2010; Siu & US Preventive Services Task Force, 2016). Consistent with this, few primary studies collect or report diagnostic status for dysthymia or persistent depressive disorder. Datasets where not all participants were eligible were included if primary data allowed selection of eligible participants. For the present study, we only included primary datasets from the HADS-D IPDMA that also provided HADS-T scores or item scores to calculate HADS-T scores.

Search strategy and study selection

A medical librarian searched Medline, Medline In-Process & Other Non-Indexed Citations and PsycINFO via OvidSP, and Web of Science via ISI Web of Knowledge from inception to October 25, 2018 using a peer-reviewed (McGowan, Sampson, Salzwedel, Cogo, Foerster, & Lefebvre, 2016) search strategy (Supplementary Methods A). We also reviewed reference lists of relevant reviews and queried contributing authors about non-published studies. Search results were uploaded into RefWorks (RefWorks-COS, Bethesda, MD, USA). After de-duplication, unique citations were uploaded into DistillerSR (Evidence Partners, Ottawa, Canada) for tracking search results.

Pairs of investigators independently reviewed titles and abstracts for eligibility. If either deemed a study potentially eligible, full-text review was done by two investigators, independently, with disagreements resolved by consensus, consulting a third investigator when necessary. Translators were consulted for languages other than those for which team members were fluent.

Data contribution, extraction, and synthesis

Authors of eligible datasets were invited to contribute de-identified primary data. We emailed corresponding authors of eligible primary studies at least three times, as necessary. If we did not receive a response, we emailed co-authors and attempted to contact corresponding authors by phone.

Diagnostic interview and country were extracted from published reports by pairs of investigators independently, with disagreements resolved by consensus. Countries were categorized as “very high”, “high” or “low-medium” development based on the United Nation’s Human Development Index (HDI) for the country for the year of the study publication. The HDI is a statistical composite index that includes indicators of life expectancy, education, and income (United Nations Development Programme, 2020). Participant-level data included age, sex, participant recruiting setting, HADS-D scores, HADS-T scores, and major depression status (case or non-case). For defining major depression, we considered MDD or MDE based on the DSM or ICD. If more than one was reported, we prioritized MDE over MDD (because screening would attempt to detect depressive episodes and further interview would determine if the episode is related to MDD, bipolar disorder or persistent depressive disorder). We also prioritized DSM over ICD because most studies use DSM criteria.

Individual participant data were converted to a standard format and synthesized into a single dataset with study-level data. We compared published participant characteristics and diagnostic accuracy estimates with results from raw datasets and resolved any discrepancies in consultation with primary study investigators.

Risk of Bias Assessment

Risk of bias of included studies was assessed by two investigators independently using the QUality Assessment of Diagnostic Accuracy Studies-2 tool (QUADAS-2; Supplementary Methods B) (Whiting et al., 2011). Any discrepancies were resolved via consensus with a third investigator involved as necessary. Risk of bias was coded at both study and participant levels since some classifications (e.g., the time between index test and reference standard) may have differed among participants from the same study. The QUADAS-2 results were used to describe the risk of bias of each included study.

Statistical Analyses

To compare the screening accuracy of the HADS-D and HADS-T across relevant cutoffs to detect major depression, we first estimated overall sensitivity and specificity for HADS-D and HADS-T by combining all studies regardless of reference standard. Reference standards used in primary studies included semi-structured interviews (e.g., SCID (First, 1995)), fully structured interviews (the Mini International Neuropsychiatric Interview (MINI) excluded) (e.g., Composite International Diagnostic Interview (CIDI) (Robins et al., 1988)), and the MINI (Lecrubier et al., 1997; Sheehan et al., 1997). Different types of reference standards have different design and performance characteristics (Levis, Benedetti, et al., 2019; Levis et al., 2020; Wu, Levis, Ioannidis, et al., 2021; Wu, Levis, Sun, et al., 2020), and estimates of sensitivity and specificity differ by type (Negeri, et al., 2021; Levis, Benedetti, et al., 2019; Levis et al., 2020; Wu, Levis, Sun, et al., 2021). It is reasonable to assume, though, that differences in sensitivity and specificity between HADS-D and HADS-T accuracy among the same participants are not associated with reference standard type, since in each primary study the HADS-D and HADS-T were compared to the same reference standard. Thus, our main analysis included all studies regardless of reference standard.

Separately, as a sensitivity analysis, to ensure that results would not differ by clinical interview, we repeated all analyses for only studies that used a semi-structured interview as the reference standard. Semi-structured interviews (e.g., SCID (First, 1995), Schedules for Clinical Assessment in Neuropsychiatry (WHO, 1994), Schedule for Affective Disorders and Schizophrenia (Endicott & Spitzer, 1987), and Monash Interview for Liaison Psychiatry (Clarke, Smith, Herrman, & McKenzie, 1998)) are intended to be administered by experienced diagnosticians and are considered to more closely reflect clinical diagnostic procedures than fully structured interviews or the MINI (Brugha, Bebbington, & Jenkins, 1999; Brugha, Jenkins, Taub, Meltzer, & Bebbington, 2001; Nosen & Woody, 2008). We did not conduct additional sensitivity analyses with fully structured interviews or the MINI.

Overall and separately, for studies that used a semi-structured reference standard, for all possible cutoffs 0-21 of the HADS-D and 0-42 of the HADS-T, we fitted bivariate random-effects models via Gauss-Hermite quadrature (Riley, Dodd, Craig, Thompson, & Williamson, 2008). This is a two-stage meta-analytic approach that models sensitivity and specificity simultaneously and accounts for the correlation between them and the precision of estimates within studies. We also constructed empirical receiver operating characteristic (ROC) plots based on pooled sensitivity and specificity estimates and calculated area under the curves (AUC) for the two tests.

To investigate heterogeneity across studies, overall and for studies with a semi-structured reference standard, we generated forest plots for the differences in sensitivity and specificity estimates between the HADS-D and HADS-T for the optimal cutoffs based on pooled results. We also quantified heterogeneity at the optimal cutoffs for the HADS-D and HADS-T by reporting the estimated variances of the random effects for the differences in the HADS-D and

HADS-T sensitivity and specificity (τ^2) (Fagerland, Lydersen, & Laake, 2014; Higgins & Thompson, 2002).

To compare the diagnostic accuracy of the HADS-D and HADS-T, using the analyses that pooled across reference standards and within semi-structured reference standard category, we first calculated the differences of the AUCs with 95% confidence intervals (CIs). Second, we compared the ROC plots visually to determine if one measure consistently perform better than the other across cutoffs. Third, we compared differences in sensitivity and specificity for optimal cutoffs and other cutoffs close to the optimal cutoff to determine if there were differences and the magnitude of any differences. To do this, we identified the optimal cutoff that minimized the values of the distance to the top-left corner of the ROC curves (NCSS, 2017) for both HADS-D and HADS-T and a set of other cutoffs that were close to the optimal cutoff. The distance to the top-left corner of the ROC curve for each cutoff value is calculated by d = $\sqrt{(\text{1-Sensitivity})^2+(\text{1-Specificity})^2}$ (NCSS, 2017). Since there is no *a priori* method to align cutoffs on the HADS-D and HADS-T that perform most similarly in terms of sensitivity and specificity, we did this based on examination of results and consensus among investigators. Then, we compared the sensitivity and specificity between the HADS-D and HADS-T for pairs of optimal cutoffs and four other pairs of cutoffs close to the optimal; the interval between cutoffs for HADS-T was 2 instead of 1 because HADS-T doubled the length and the total score of HADS-D. For all cutoffs on the HADS-D and HADS-T, 95% CIs for the differences between HADS-D and HADS-T sensitivity and specificity were constructed via a cluster bootstrap approach (Van der Leeden, Busing, & Meijer, 1997; Van der Leeden, Meijer, & Busing, 2008) with resampling at the study and subject level. For each comparison, we ran 1000 iterations of the bootstrap. For each bootstrap iteration, the bivariate random-effects model was fitted to the

HADS-D and HADS-T data, and the pooled sensitivities and specificities were computed separately, as described above, for all cutoffs of HADS-D and HADS-T.

In addition to comparing the HADS-D and HADS-T with pooling of study-level results, as a sensitivity analysis, we compared sensitivity and specificity of the HADS-D and HADS-T across cutoffs via an individual-level analysis. For the individual-level analysis, for each pair of matched HADS-D and HADS-T cutoffs, we fitted a linear mixed model with the difference between the HADS-D and HADS-T screening results as the outcome. The screening result is dichotomous, either positive = 1 or negative = 0. If the HADS-T screening result was positive (which was 1), but HADS-D was negative (which was 0), the outcome, i.e., the difference between HADS-T and HADS-D results, was 1 − 0 = 1; if both screening results were positive or negative, the outcome was 0 (1 – 1 or 0 − 0); and if the HADS-T screening result was negative, but HADS-D was positive, the outcome was -1 (0 – 1 = -1). This model modeled the differences in sensitivity and specificity simultaneously and included random effects both at the study level. From this model, for each set of HADS-D and HADS-T paired cutoffs, we estimated the difference in sensitivity and specificity between the two tests and associated CIs. These CIs from the bootstrap approach and individual-level analysis allowed us to test whether the sensitivity and specificity of the HADS-T is equivalent to that of the HADS-D based on a pre-specified equivalence margin of δ = 0.05 (Walker & Nowacki, 2011), as we have done in previous studies (Harel et al., 2021; Ishihara et al., 2019; Wu, Levis, Riehm, et al., 2020).

As a sensitivity analysis, we compared accuracy of HADS-D and HADS-T results stratified by subgroups based on inpatient and outpatient care settings (we planned to conduct sensitivity analysis in each participant recruit setting, separately, but we were able to do this only for inpatient and outpatient medical settings because there were too few participants from non-medical and mixed inpatient/outpatient settings). In addition, we conducted a subgroup analysis only among patients from cancer studies because meta-analyses (Mitchell et al., 2010; Vodermaier & Millman, 2011) of studies from cancer care settings reported that the HADS-T may perform better than the HADS-D in those settings. We did not conduct the sensitivity analysis to assess whether inclusion of published results from the eligible studies that did not provide raw data influenced results because we did this in the main HADS-D IPDMA and found no differences (Wu et al., 2021).

To examine whether measurement differences across participant characteristics, including country, may have influenced our results, we assessed whether sensitivity and specificity differed for the HADS-D based on these characteristics, and then, we re-examined HADS-D and HADS-T differences for any variables where differences were found. To assess possible influences on sensitivity and specificity, we conducted one-stage meta-regressions. In the first step, we repeated the analysis that we did in the main HADS-D IPDMA by interacting all subgrouping variables (age [measured continuously], sex [reference category = female]), country HDI level [reference category = very high], cancer diagnosis [reference category = no], participant recruiting setting [reference category = inpatient specialty care], interactions of QUADAS-2 signaling item responses [reference category = low risk] with logit (sensitivity) and logit (1 – specificity) of the HADS-D (Wu et al., 2021). We conducted these analyses separately by reference standards (semi-structured interview, fully structured interview, MINI), since these types of interviews have been shown to identify different individuals (Wu et al., 2021). In the second step, we added country/language variables to the model (Germany, Spain, Lithuania, Norway, Korea, Japan [reference category = English speaking countries]). These models were restricted to the subset of the studies from countries with more than 500 participants that had complete data for all relevant variables and used a semi-structured interview or the MINI (there were not enough data for the studies that used a fully structured reference standard). Country HDI level was dropped from the model because all countries included in this analysis had very high HDI. For any variables that were found to be associated with the sensitivity or specificity across all cutoffs, we compared accuracy of HADS-D and HADS-T results stratified by subgroups based on these variables.

All analyses were run in R (R version R 3.5.0 (R Core Team, 2020) and R Studio version 1.1.423 (RStudio Team, 2020)) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015).

Registration and Protocol

The main HADS-D IPDMA was registered in PROSPERO (CRD42015016761), and a protocol was published (Thombs et al., 2016). The present study was not included in the protocol for the main HADS-D IPDMA, but a separate protocol was developed and posted online prior to initiating the study (https://osf.io/438ak/).

Data Availability

Data contribution agreements with primary study authors do not include permission to make their data publicly available, although the dataset used in this study will be archived through a McGill University repository (Borealis, https://borealisdata.ca/dataverse/depressdproject/). The R codes used for the analysis will be made publicly available through the same repository. Requests to access the dataset to verify study results but not for other purposes can be sent to the corresponding authors via the "Access Dataset" function on the repository website.

Results

Search Results and Inclusion of Primary Data

For the main HADS-D IPDMA, of 14,465 unique titles and abstracts identified from the database search, 13,895 were excluded after title and abstract review and 330 after full-text (Supplementary Table A), leaving 240 eligible articles with data from 165 unique participant samples (Supplementary Figure A). Of the 165 unique samples, 93 (56%) contributed data (66% of eligible participants). In addition, authors of included studies contributed data from 10 studies that were unpublished or did not come up in the search, for a total of 103 HADS-D datasets contributed to our IPDMA. Five studies without HADS individual item scores or separate total scores for the HADS-D and HADS-T were excluded from the present study (see Supplementary Table B2). Thus, 20,700 participants (2,285 major depression cases) from 98 studies were analyzed (91% of 22,755 participants from the 103 HADS-D datasets). Included study characteristics are shown in Supplementary Table B1. Characteristics of eligible studies that did not provide data, including the five studies excluded because they only provided HADS-D or HADS-T total scores, are shown in Supplementary Table B2.

Of 98 included studies, 58 used semi-structured interviews to assess major depression (10,311 participants), including 54 that used the SCID (9,676 participants); 31 used the MINI (7,445 participants); and 9 used other. Participant characteristics are shown in Table 1.

Supplementary Table C shows QUADAS-2 ratings for included studies. There were only 11 studies with "low" risk of bias rating across all QUADAS-2 domains.

Comparison of Screening Accuracy Between the HADS-D and HADS-T

ROC plots comparing sensitivity and specificity estimates for all cutoffs between the HADS-D (0-21) and HADS-T (0-42) among all included studies are shown in Figure 1. A large part of the plots for the HADS-D and HADS-T were overlapping. The HADS-T performed better than HADS-D at some cutoffs, but this pattern was not consistent across cutoffs. The AUCs for the HADS-D and HADS-T were similar among all studies (0.853 versus 0.872). We also compared the ROCs among studies that used a semi-structured reference standard and found a similar pattern (Supplementary Figure B).

Based on the pooled sensitivity and specificity across all HADS-D and HADS-T cutoffs, among all studies, the cutoff that minimized the values of the distance to the top-left corner of the ROC curves was ≥ 7 for the HADS-D (sensitivity [95% CI] = 0.79 [0.75, 0.83], specificity [95% CI] = 0.78 [0.75, 0.80]) and ≥ 15 for the HADS-T (sensitivity [95% CI] = 0.79 [0.76, 0.82], specificity [95% CI] = 0.81 [0.78, 0.83]) (Table 2).

The comparison of sensitivity and specificity between the HADS-D and HADS-T for the optimal cutoffs (HADS-D ≥ 7 vs. HADS-T ≥ 15) and other cutoffs close to the optimal cutoffs (≥ 5 vs. ≥ 11; ≥ 6 vs. ≥ 13; ≥ 8 vs. ≥ 17; ≥ 9 vs. ≥ 19; ≥ 10 vs. ≥ 21; and ≥ 11 vs. ≥ 23 are presented in Table 2. Overall, for the pairs of optimal cutoffs or other cutoffs close to the optimal, the differences in sensitivity and specificity between HADS-D and HADS-T using the bootstrapping approach across all 98 primary studies were small. Precision of estimates was high, and the width of 95% CIs ranged from 5% to 9% for sensitivity and 2% to 4% for specificity across all cutoffs examined. For sensitivity, the differences of HADS-T − HADS-D for all pairs of cutoffs were not statistically significant (the differences were between -0.05 and 0.01, CIs were within or overlapped with the range of -0.05 and 0.05). Therefore, at five pairs of optimal cutoffs or other cutoffs close to the optimal, the sensitivity of the HADS-T was equivalent to that of the HADS-D; the equivalency was indeterminant on the other two pairs, based on the pre-specified equivalence margin of δ = 0.05. For specificity, estimates of HADS-T were equivalent to HADS-D for all seven pairs of cutoffs (the differences of HADS-T − HADS-D were between 0.02 and

0.03; CIs were all within -0.05 and 0.05). Relevant results among studies that used a semi-structured reference standard were consistent with overall estimates (Supplementary Table D1).

The comparison of results via individual-level analysis are presented in Table 3. For each pair of matched HADS-D and HADS-T cutoffs, the differences in sensitivity and specificity between the two tests were similar to those from the bivariate random-effects models. This was also true among studies that used a semi-structured reference standard (Supplementary Table D2).

Among participants in inpatient care settings (Table 4a; 8,827 participants from 38 studies), the comparison results of HADS-T − HADS-D in sensitivity were similar to the overall estimates; the differences in specificity were slightly larger than overall estimates, however, the 95% CIs generally overlapped with -0.05 and 0.05 and were classified as indeterminate to equivalency, with one exception (HADS-D ≥ 6 vs. HADS-T ≥ 13) for which HADS-T specificity was greater than for the HADS-D. The comparison results among participants in outpatient care settings (Table 4b; 9,547 participants from 54 studies) and participants from studies done in cancer care settings (Supplementary Table E; 5608 participants from 23 studies) were similar to overall estimates. Within the semi-structured reference standard category, similar patterns were found (Supplementary Tables D3 and D4).

The meta-regression results indicated no significant differences in sensitivity and specificity were found for any individual participant characteristics or risk of bias ratings (Supplementary Table F1-F3). After adding the country/language variables to the model, the sensitivity and specificity of HADS-D was invariant based on all variables across reference standards except that specificity estimates of the HADS-D were associated with Germany and Spain among studies that used a semi-structured reference standard; specifically, the HADS-D had lower specificity among participants from Germany and Spain compared to studies done with participants from English speaking countries (Supplementary Table G1-G2).

Therefore, we conducted subgroup analysis of our comparisons of HADS-D and HADS-T accuracy for participants from Germany or Spain. For each pair of matched HADS-D and HADS-T cutoffs among participants from Germany (Supplementary Table H1), the comparison results of HADS-T − HADS-D in sensitivity and specificity were similar to the overall estimates; among participants from Spain (Supplementary Table H2), differences in specificity were slightly larger than overall estimates, however, the 95% CIs all overlapped with -0.05 and 0.05 and were classified as indeterminate to equivalent, and differences in sensitivity were similar to the overall estimates.

A forest plot of the differences of sensitivity and specificity estimates for HADS-D ≥ 7 vs. HADS-T ≥ 15 across all studies is shown in Figure 2. At the optimal cutoffs, there was low heterogeneity in the differences between HADS-D and HADS-T across the 98 studies with estimated inter-study heterogeneity (τ^2) < 0.01 for sensitivity and < 0.01 for specificity. The forest plot of the differences of sensitivity and specificity estimates at optimal cutoffs for the HADS-D and HADS-T among studies that used a semi-structured reference standard is shown in Supplementary Figure C.

Discussion

We assessed the equivalency of screening accuracy of the HADS-D and HADS-T across all cutoffs to detect major depression and compared accuracy across paired optimal cutoffs and other cutoffs close to the optimal cutoffs to test whether the HADS-T is superior to HADS-D for major depression detection. There were two main findings. First, among all 98 included studies the values of the distance to the top-left corner of the ROC curves (Riley et al., 2008) were minimized at a HADS-D cutoff ≥ 7 (sensitivity = 0.79, specificity = 0.78) and at a HADS-T cutoff ≥ 15 (sensitivity = 0.79, specificity = 0.81). Second, at paired optimal cutoffs and six other cutoffs close to the optimal cutoffs, the HADS-D was similarly accurate compared to the HADS-T overall and among studies that used a semi-structured reference standard.

Overall, for all 98 primary studies, across all sets of paired cutoffs, the sensitivity and specificity of the HADS-T were classified as equivalent to that of the HADS-D based on the pre-specified equivalency margin. Although the HADS-T was slightly more specific (range 0.02 to 0.03), all the 95% CIs for differences in sensitivity and specificity of HADS-T − HADS-D were within or overlapped with the range of -0.05 and 0.05. When we analyzed data separately among studies that used a semi-structured reference standard, differences in sensitivity and specificity between the HADS-D and HADS-T were similar to the overall estimates.

Furthermore, similar to overall estimates, there were no substantive differences in performance between the HADS-D and HADS-T in detecting major depression among medical outpatients. Among inpatients, the HADS-T and HADS-D were also equivalent in sensitivity. The HADS-T performed slightly better than HADS-D in terms of specificity, and equivalency was indeterminant based on the pre-specified equivalence margin, except for one pair of cutoffs. This finding is possibly related to the greater presence of anxiety symptoms in inpatients versus outpatients and its relationship to depression (Schatzberg, 2019).

Previous conventional meta-analyses of results from cancer patients (Mitchell et al., 2010; Vodermaier & Millman, 2011) suggested that the HADS-T may perform better than the HADS-D, but that conclusion was highly uncertain given the limitations of the samples and methods. Through our IPDMA, with its large dataset and more rigorous comparison methods including both bivariate random-effects models and individual-level models, a two-level bootstrap approach (Fagerland et al., 2014; Higgins & Thompson, 2002), and subgroup analysis, we found there was no consistent evidence that the HADS-T is superior to HADS-D for major depression detection, including in cancer care settings. In addition, we did not identify any differences between HADS-D and HADS-T accuracy that were associated with individual participant characteristics or countries. Therefore, in research and clinical general practice, using the full 14-item HADS-T for depression screening would likely result in no to minimal gain in screening accuracy but would add unnecessary burden to patients compared to the 7-item HADS-D.

To our knowledge, this is the first meta-analysis that directly compared the HADS-D and HADS-T for screening for depression using the same large individual participant dataset for both screening tools. Strengths of this study included the large overall sample size and high precision of estimates of differences, the ability to compare results for HADS-D and HADS-T across all cutoffs from all studies, and the ability to assess screening accuracy overall and by inpatient and outpatient subgroups. There are also limitations to consider. First, for the full IPDMA data, primary data from 72 of 165 published eligible datasets (44% of datasets, 34% of participants) were not included, and only those datasets with complete data for all individual HADS item scores (91% of available data) were included in this study. Nonetheless, this sample was much larger than the few primary studies that have previously compared the HADS-D and HADS-T. Second, we did not conduct analyses restricted to studies with “low” risk of bias ratings across QUADAS-2 domains. However, in sensitivity analysis in this study and in our main IPDMA on the HADS-D (Wu, et al., 2021), risk of bias ratings were not associated with screening accuracy. Third, the present study used a subset of studies and participants from our previously conducted HADS-D IPDMA (Wu, et al., 2021). This IPDMA project was designed to assess the accuracy of the HADS-D for detecting major depression. Diagnoses of other mental disorders, including, anxiety disorders, were not collected in most of the included primary studies. Thus, we were not able to evaluate the sensitivity and specificity of the HADS-D, HADS-Anxiety, or HADS-T for detecting mental disorders generally. Forth, we did not record inter-rated reliability for risk of bias ratings; however, all ratings were done by trained reviewers and any disagreements were addressed by consensus, including a third investigator as necessary.

Conclusions

In summary, this study found that sensitivity and specificity of the HADS-T were not superior to the HADS-D for detecting major depression in a large individual participant dataset. Using the 7-item HADS-D for depression screening instead of the full 14-item HADS-T has minimal influence on performance of the measure but would reduce patient and participant burden in clinical and research settings. Both HADS-D and HADS-T have only modest screening ability and discussion of their exact indications for use and related caveats are beyond the scope of this article. However, there were no substantive differences in performance between the HADS-D and HADS-T in detecting major depression among medical outpatients, although there was a slight advantage in specificity of indeterminate equivalency for the HADS-T among medical inpatients, for whom adding the anxiety items of HADS-A may improve accuracy.

Ethical Approval: As this study involved secondary analysis of anonymized previously collected data, the Research Ethics Committee of the Jewish General Hospital declared that this project did not require research ethics approval. However, for each included dataset, we confirmed that the original study received ethics approval and that all patients provided informed consent.

REFERENCES

*Akechi, T., Okuyama, T., Sugawara, Y., Shima, Y., Furukawa, T. A., & Uchitomi, Y. (2006). Screening for depression in terminally ill cancer patients in Japan. *Journal of Pain & Symptom Management* , 31(1), 5-12.

*Al-Asmi, A., Dorvlo, A. S., Burke, D. T., Al-Adawi, S., Al-Zaabi, A., Al-Zadjali, H. A., … & Al-Adawi, S. (2012). The detection of mood and anxiety in people with epilepsy using two-phase designs: experiences from a tertiary care centre in Oman. *Epilepsy Research, 98*(2-3), 174–181.

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders: DSM-III. 3rd ed. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (1994) Diagnostic and statistical manual of mental disorders: DSM-IV. 4th ed. Washington (DC): American Psychiatric Association.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders: DSM-IV-TR. 4th ed, Text Revision. Washington (DC): American Psychiatric Association.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Washington (DC): American Psychiatric Association.

*Amoozegar, F., Patten, S. B., Becker, W. J., Bulloch, A. G., Fiest, K. M., Davenport, W. J., ... & Jette, N. (2017). The prevalence of depression and the accuracy of depression screening tools in migraine patients. *General Hospital Psychiatry*, *48*, 25-31.

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01

*Bayón-Pérez, C., Hernando, A., Álvarez-Comino, M., Cebolla, S., Serrano, L., Gutiérrez, F., … & Pulido, F. (2016). Toward a comprehensive care of HIV patients: finding a strategy to detect depression in a Spanish HIV cohort. *AIDS Care, 28*, 834 - 841.

*Beck, K. R., Tan, S. M., Lum, S. S., Lim, L. E., & Krishna, L. K. (2016). Validation of the emotion thermometers and hospital anxiety and depression scales in Singapore: Screening cancer patients for distress, anxiety and depression. *Asia-Pacific Journal of Clinical Oncology*, *12*(2), e241–e249.

*Beraldi, A., Baklayan, A., Hoster, E., Hiddemann, W., & Heussner, P. (2014). Which questionnaire is most suitable for the detection of depressive disorders in haemato-oncological patients? Comparison between HADS, CES-D and PHQ-9. *Oncology Research and Treatment, 37*,108.

*Bernstein, C. N., Zhang, L., Lix, L. M., Graff, L. A., Walker, J. R., Fisk, J. D., ... & CIHR Team in Defining the Burden and Managing the Effects of Immune-mediated Inflammatory Disease. (2018). The validity and reliability of screening measures for depression and anxiety disorders in inflammatory bowel disease. *Inflammatory Bowel Diseases*, *24*(9), 1867-1875.

*Braeken, A. P. B. M., Lechner, L., Houben, R. M. A., Van Gils, F. C. J. M., & Kempen, G. I. J. M. (2011). Psychometric properties of the Screening Inventory of Psychosocial Problems (SIPP) in Dutch cancer patients treated with radiotherapy. *European Journal of Cancer Care*, 20(3), 305-314.

Brugha, T. S., Bebbington, P. E., & Jenkins, R. (1999). A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychological Medicine, 29*(5), 1013-1020. doi:10.1017/S0033291799008880

Brugha, T. S., Jenkins, R., Taub, N., Meltzer, H., & Bebbington, P. E. (2001). A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine, 31*(6), 1001-1013. doi:10.1017/S0033291701004184

*Bunevicius, A., Peceliuniene, J., Mickuviene, N., Valius, L., & Bunevicius, R. (2007). Screening for depression and anxiety disorders in primary care patients. *Depression and Anxiety, 24*(7), 455–460.

*Bunevicius, A., Staniute, M., Brozaitiene, J., & Bunevicius, R. (2012). Diagnostic accuracy of self-rating scales for screening of depression in coronary artery disease patients. *Journal of Psychosomatic Research, 72*(1), 22–25.

*Butnoriene, J., Bunevicius, A., Norkus, A., & Bunevicius, R. (2014). Depression but not anxiety is associated with metabolic syndrome in primary care based community sample. *Psychoneuroendocrinology, 40*, 269–276.

*Can, C., Cimilli, C., Ozenli, Y., Ergor, G., Aysevener, E. O., Unek, T., & Astarcioglu, I. (2018). Quality of life and psychiatric disorders before and one year after liver transplantation. *Journal of Clinical and Analytical Medicine, 9*(5), 396-401.

*Chen, C. K., Tsai, Y. C., Hsu, H. J., Wu, I. W., Sun, C. Y., Chou, C. C., … & Wang, L. J. (2010). Depression and suicide risk in hemodialysis patients with chronic renal failure. *Psychosomatics, 51*(6), 528–528.e6.

*Cheung, G., Patrick, C., Sullivan, G., Cooray, M., & Chang, C. L. (2012). Sensitivity and specificity of the Geriatric Anxiety Inventory and the Hospital Anxiety and Depression

Scale in the detection of anxiety disorders in older people with chronic obstructive pulmonary disease. *International Psychogeriatrics*, *24*(1), 128–136.

Clarke, D. M., Smith, G. C., Herrman, H. E., & McKenzie, D. P. (1998). Monash Interview for Liaison Psychiatry (MILP) - Development, reliability, and procedural validity. *Psychosomatics, 39*(4), 318-328. doi:10.1016/S0033-3182(98)71320-9

*Consoli, S. M., Rolhion, S., Martin, C., Ruel, K., Cambazard, F., Pellet, J., & Misery, L. (2006). Low levels of emotional awareness predict a better response to dermatological treatment in patients with psoriasis. *Dermatology, 212*(2), 128–136.

*Costa-Requena, G., Ballester Arnal, R., & Gil, F. (2013). Perceived social support in Spanish cancer outpatients with psychiatric disorder. *Stress and Health: Journal of the International Society for the Investigation of Stress, 29*(5), 421–426.

*Cukor, D., Coplan, J., Brown, C., Friedman, S., Newville, H., Safier, M., ... & Kimmel, P. L. (2008). Anxiety disorders in adults treated by hemodialysis: a single-center study. *American Journal of Kidney Diseases, 52*(1), 128-136.

*da Rocha e Silva, C. E. D. R., Brasil, M. A. A., Do Nascimento, E. M., de Bragança Pereira, B., & André, C. (2013). Is poststroke depression a major depression?. *Cerebrovascular Diseases*, 35(4), 385-391.

*De Souza, J., Jones, L. A., & Rickards, H. (2010). Validation of self-report depression rating scales in Huntington's disease. *Movement Disorders*, 25(1), 91-96.

*de la Torre, A. Y., Oliva, N., Echevarrieta, P. L., Pérez, B. G., Caporusso, G. B., Titaro, A. J., … & Daray, F. M. (2016). Major depression in hospitalized Argentine general medical patients: Prevalence and risk factors. *Journal of Affective Disorders*, *197,* 36–42.

*de Oliveira, G. N., Lessa, J. M., Gonçalves, A. P., Portela, E. J., Sander, J. W., & Teixeira, A. L. (2014). Screening for depression in people with epilepsy: comparative study among neurological disorders depression inventory for epilepsy (NDDI-E), hospital anxiety and depression scale depression subscale (HADS-D), and Beck depression inventory (BDI). *Epilepsy & Behavior, 34,* 50–54.

*Douven, E., Schievink, S. H., Verhey, F. R., van Oostenbrugge, R. J., Aalten, P., Staals, J., & Köhler, S. (2016). The Cognition and Affect after Stroke - a Prospective Evaluation of Risks (CASPER) study: rationale and design. *BMC Neurology, 16*, 65.

*Dorow, M., Stein, J., Pabst, A., Weyerer, S., Werle, J., Maier, W., ... & Riedel-Heller, S. G. (2018). Categorical and dimensional perspectives on depression in elderly primary care patients–Results of the AgeMooDe study. *International Journal of Methods in Psychiatric Research*, 27(1), e1577.

*Drabe, N., Zwahlen, D., Büchi, S., Moergeli, H., Zwahlen, R. A., & Jenewein, J. (2008). Psychiatric morbidity and quality of life in wives of men with long-term head and neck cancer. *Psycho-Oncology, 17*(2), 199–204.

Endicott, J., & Spitzer, R. L. (1987). [Schedule for Affective Disorders and Schizophrenia (SADS)]. *Acta psychiatrica Belgica, 87*(4), 361-516.

*Fábregas, B. C., Moura, A. S., Ávila, R. E., Faria, M. N., Carmo, R. A., & Teixeira, A. L. (2014). Sexual dysfunction and dissatisfaction in chronic hepatitis C patients. *Revista da Sociedade Brasileira de Medicina Tropical, 47*(5), 564–572.

Fagerland, M. W., Lydersen, S., & Laake, P. (2014). Recommended tests and confidence intervals for paired binomial proportions. *Statistics in Medicine, 33*(16), 2850-2875. doi:10.1002/sim.6148

*Ferentinos, P., Paparrigopoulos, T., Rentzos, M., Zouvelou, V., Alexakis, T., & Evdokimidis, I. (2011). Prevalence of major depression in ALS: comparison of a semi-structured interview and four self-report measures. *Amyotrophic Lateral Sclerosis*, 12(4), 297-302..

*Fiest, K. M., Patten, S. B., Wiebe, S., Bulloch, A. G., Maxwell, C. J., Jette, N. (2014). Validating screening tools for depression in epilepsy. *Epilepsia, 55*(10), 1642-1650.

First, M. B. (1995). Structured Clinical Interview for the DSM (SCID). New York (NY): John Wiley & Sons, Inc.

*Fischer, H. F., Klug, C., Roeper, K., Blozik, E., Edelmann, F., Eisele, M., … & Herrmann-Lingen, C. (2014). Screening for mental disorders in heart failure patients using computer-adaptive tests. *Quality of Life Research, 23*(5), 1609–1618.

*Gagnon, N., Flint, A. J., Naglie, G., & Devins, G. M. (2005). Affective correlates of fear of falling in elderly persons. *The American Journal of Geriatric Psychiatry, 13*(1), 7–14.

*Gandy, M., Sharpe, L., Perry, K. N., Miller, L., Thayer, Z., Boserio, J., & Mohamed, A. (2012). Assessing the efficacy of 2 screening measures for depression in people with epilepsy. *Neurology, 79*(4), 371–375.

*Golden, J., Conroy, R. M., & O'Dwyer, A. M. (2007). Reliability and validity of the Hospital Anxiety and Depression Scale and the Beck Depression Inventory (Full and FastScreen scales) in detecting depression in persons with hepatitis C. *Journal of Affective Disorders, 100*(1-3), 265–269.

*Gould, K. R., Ponsford, J. L., Johnston, L., & Schönberger, M. (2011). Predictive and associated factors of psychiatric disorders after traumatic brain injury: a prospective study. *Journal of Neurotrauma, 28*(7), 1155–1163.

*Grassi, L., Sabato, S., Rossi, E., Marmai, L., & Biancosino, B. (2009). Affective syndromes and their screening in cancer patients with early and stable disease: Italian ICD-10 data and performance of the Distress Thermometer from the Southern European Psycho-Oncology Study (SEPOS). *Journal of Affective Disorders*, *114*(1-3), 193–199.

*Hahn, D., Reuter, K., & Härter, M. (2006). Screening for affective and anxiety disorders in medical patients - comparison of HADS, GHQ-12 and Brief-PHQ. *GMS Psycho-Social Medicine, 3*.

Harel, D., Levis, B., Ishihara, M., Levis, A. W., Vigod, S. N., Howard, L. M., . . . DEPRESsion Screening Data (DEPRESSD) EPDS Collaboration. (2021). Shortening the Edinburgh postnatal depression scale using optimal test assembly methods: Development of the EPDS-Dep-5. *Acta Psychiatrica Scandinavica*, *143*(4), 348-362. doi:10.1111/acps.13272

*Härter, M., Woll, S., Wunsch, A., Bengel, J., & Reuter, K. (2005). Screening for mental disorders in cancer, cardiovascular and musculoskeletal diseases. *Social Psychiatry and Psychiatric Epidemiology, 41*, 56-62.

*Hartung, T. J., Friedrich, M., Johansen, C., Wittchen, H. U., Faller, H., Koch, U., Brähler, E., … & Mehnert, A. (2017). The Hospital Anxiety and Depression Scale (HADS) and the 9-item Patient Health Questionnaire (PHQ-9) as screening instruments for depression in patients with cancer. *Cancer*, *123*(21), 4236–4243.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics In Medicine*, *21*(11), 1539-1558. doi:10.1002/sim.1186

*Hitchon, C. A., Zhang, L., Peschken, C. A., Lix, L. M., Graff, L. A., Fisk, J. D., Patten, S. B., Bolton, J., Sareen, J., El-Gabalawy, R., Marriott, J., Bernstein, C. N., & Marrie, R. A.

(2020). Validity and Reliability of Screening Measures for Depression and Anxiety Disorders in Rheumatoid Arthritis. *Arthritis Care & Research, 72*(8), 1130–1139.

*Honarmand, K., & Feinstein, A. (2009). Validation of the Hospital Anxiety and Depression Scale for use with multiple sclerosis patients. *Multiple Sclerosis, 15*(12), 1518–1524.

*Huey, N. S., Guan, N. C., Gill, J. S., Hui, K. O., Sulaiman, A. H., & Kunagasundram, S. (2018). Core Symptoms of Major Depressive Disorder among Palliative Care Patients. *International Journal of Environmental Research and Public Health, 15*(8), 1758.

Ishihara, M., Harel, D., Levis, B., Levis, A. W., Riehm, K. E., Saadat, N., . . . Thombs, B. D. (2019). Shortening self-report mental health symptom measures through optimal test assembly methods: Development and validation of the Patient Health Questionnaire-Depression-4. *Depression and Anxiety, 36*(1), 82-92. doi:10.1002/da.22841

*Jackson, M. L., Tolson, J., Schembri, R., Bartlett, D., Rayner, G., Lee, V. V., & Barnes, M. (2021). Does continuous positive airways pressure treatment improve clinical depression in obstructive sleep apnea? A randomized wait-list controlled study. *Depression and Anxiety, 38*(5), 498–507.

*Jang, J. E., Kim, S. W., Kim, S. Y., Kim, J. M., Park, M. H., Yoon, J. H., … & Yoon, J. S. (2013). Religiosity, depression, and quality of life in Korean patients with breast cancer: a 1-year prospective longitudinal study. *Psycho-Oncology, 22*(4), 922–929.

Joffres, M., Jaramillo, A., Dickinson, J., Lewin, G., Pottie, K., Shaw, E., Gorber, S. C., & Tonelli, M. (2013). Recommendations on screening for depression in adults. *CMAJ : Canadian Medical Association Journal, 185*(9), 775–782. https://doi.org/10.1503/cmaj.130403

*Juliao, M., Barbosa, A., Oliveira, F., & Nunes, B. (2013). Prevalence and factors associated with desire for death in patients with advanced disease: results from a Portuguese cross-sectional study. *Psychosomatics, 54*(5), 451–457.

*Kang, H. J., Stewart, R., Kim, J. M., Jang, J. E., Kim, S. Y., Bae, K. Y., … & Yoon, J. S. (2013). Comparative validity of depression assessment scales for screening poststroke depression. *Journal of Affective Disorders*, *147*(1-3), 186–191.

*Keller, M., Sommerfeldt, S., Fischer, C., Knight, L., Riesbeck, M., Lowe, B., … & Lehnert, T. (2004). Recognition of distress and psychiatric morbidity in cancer patients: a multi-method approach. *Annals of Oncology*, *15*(8),1243.

*Kjaergaard, M., Arfwedson Wang, C. E., Waterloo, K., & Jorde, R. (2014). A study of the psychometric properties of the Beck Depression Inventory-II, the Montgomery and Åsberg Depression Rating Scale, and the Hospital Anxiety and Depression Scale in a sample from a healthy population. *Scandinavian Journal of Psychology, 55*(1), 83–89.

*Kugaya, A., Akechi, T., Okuyama, T., Nakano, T., Mikami, I., Okamura, H., & Uchitomi, Y. (2000). Prevalence, predictive factors, and screening for psychologic distress in patients with newly diagnosed head and neck cancer. *Cancer, 88*(12), 2817–2823.

*Lambert, S. D., Clover, K., Pallant, J. F., Britton, B., King, M. T., Mitchell, A. J., Carter, G. (2015). Making Sense of Variations in Prevalence Estimates of Depression in Cancer: A Co-Calibration of Commonly Used Depression Scales Using Rasch Analysis. *Journal of the National Comprehensive Cancer Network, 13*(10),1203.

*Law, M., Naughton, M. T., Dhar, A., Barton, D., & Dabscheck, E. (2014). Validation of two depression screening instruments in a sleep disorders clinic. *Journal of Clinical Sleep Medicine, 10*(6), 683–688.

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., . . . Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry, 12*(5), 224-231. doi:10.1016/S0924-9338(97)83296-8

*Lee, Y., Wu, Y. S., Chien, C. Y., Fang, F. M., & Hung, C. F. (2016). Use of the Hospital Anxiety and Depression Scale and the Taiwanese Depression Questionnaire for screening depression in head and neck cancer patients in Taiwan. *Neuropsychiatric Disease and Treatment, 12*, 2649–2657.

*Lee, C. Y., Lee, Y., Wang, L. J., Chien, C. Y., Fang, F. M., & Lin, P. Y. (2017). Depression, anxiety, quality of life, and predictors of depressive disorders in caregivers of patients with head and neck cancer: A six-month follow-up study. *Journal of Psychosomatic Research, 100*, 29–34.

*Lees, R.A., Stott, D.J., Quinn, T.J., & Broomfield, N.M. (2014). Feasibility and Diagnostic Accuracy of Early Mood Screening to Diagnose Persisting Clinical Depression/Anxiety Disorder after Stroke. *Cerebrovascular Diseases, 37*, 323 - 329.

Levis, B., Benedetti, A., Levis, A. W., Ioannidis, J. P. A., Shrier, I., Cuijpers, P., . . . Thombs, B. D. (2017). Selective Cutoff Reporting in Studies of Diagnostic Test Accuracy: A Comparison of Conventional and Individual-Patient-Data Meta-Analyses of the Patient Health Questionnaire-9 Depression Screening Tool. *American Journal of Epidemiology, 185*(10), 954-964. doi:10.1093/aje/kww191

Levis, B., Benedetti, A., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., . . . Thombs, B. D. (2018). Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. *British Journal of Psychiatry, 212*(6), 377-385. doi:10.1192/bjp.2018.54

Levis, B., Benedetti, A., Thombs, B. D., Akena, D. H., Arroll, B., Ayalon, L., . . . DEPRESsion Screening Data (DEPRESSD) Collaboration. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ, 365*. doi:10.1136/bmj.l1476

Levis, B., McMillan, D., Sun, Y., He, C., Rice, D. B., Krishnan, A., . . . Thombs, B. D. (2019). Comparison of major depression diagnostic classification probability using the SCID, CIDI, and MINI diagnostic interviews among women in pregnancy or postpartum: An individual participant data meta-analysis. *International Journal of Methods in Psychiatric Research*, *28*(4). doi:10.1002/mpr.1803

Levis, B., Negeri, Z., Sun, Y., Benedetti, A., Thombs, B. D., & DEPRESsion Screening Data (DEPRESSD) EPDS Group. (2020). Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. *BMJ, 371*. doi:10.1136/bmj.m4022

*Loosman, W. L., Siegert, C. E., Korzec, A., & Honig, A. (2010). Validity of the Hospital Anxiety and Depression Scale and the Beck Depression Inventory for use in end-stage renal disease patients. *The British Journal of Clinical Psychology, 49*(Pt 4), 507–516.

*Love, A. W., Kissane, D. W., Bloch, S., & Clarke, D. (2002). Diagnostic efficiency of the Hospital Anxiety and Depression Scale in women with early stage breast cancer. *The Australian and New Zealand Journal of Psychiatry, 36*(2), 246–250.

*Love, A. W., Grabsch, B., Clarke, D. M., Bloch, S., & Kissane, D. W. (2004). Screening for depression in women with metastatic breast cancer: a comparison of the Beck Depression Inventory Short Form and the Hospital Anxiety and Depression Scale. *The Australian and New Zealand journal of psychiatry, 38*(7), 526–531.

*Lowe, B., Gräfe, K., Zipfel, S., Spitzer, R. L., Herrmann-Lingen, C., Witte, S., & Herzog, W. (2003). Detecting panic disorder in medical and psychosomatic outpatients: comparative validation of the Hospital Anxiety and Depression Scale, the Patient Health Questionnaire, a screening question, and physicians' diagnosis. *Journal of Psychosomatic Research, 55*(6), 515–519.

*Marrie, R. A., Zhang, L., Lix, L. M., Graff, L. A., Walker, J. R., Fisk, J. D., … & Bernstein, C. N. (2018). The validity and reliability of screening measures for depression and anxiety disorders in multiple sclerosis. *Multiple Sclerosis and Related Disorders, 20*, 9–15.

*Massardo, L., Bravo-Zehnder, M., Calderón, J., Flores, P., Padilla, O., Aguirre, J. M., … & González, A. (2015). Anti-N-methyl-D-aspartate receptor and anti-ribosomal-P autoantibodies contribute to cognitive dysfunction in systemic lupus erythematosus. *Lupus, 24*(6), 558–568.

*Matsuoka, Y., Nishi, D., Nakajima, S., Yonemoto, N., Hashimoto, K., Noguchi, H., … & Kim, Y. (2009). The Tachikawa cohort of motor vehicle accident study investigating psychological distress: Design, methods and cohort profiles. *Social Psychiatry and Psychiatric Epidemiology, 44*(4), 333–340.

*McFarlane, A. C., Browne, D., Bryant, R. A., O'Donnell, M., Silove, D., Creamer, M., & Horsley, K. (2009). A longitudinal analysis of alcohol consumption and the risk of posttraumatic symptoms. *Journal of Affective Disorders, 118*(1-3), 166–172.

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.

*Meyer, A., Wollbrück, D., Täschner, R., Singer, S., Ehrensperger, C., Danker, H., … & Schwarz, R. (2008). Psychological status and morbidity of the spouses of laryngectomy patients. *Zeitschrift fur Klinische Psychologie und Psychotherapie: Forschung und Praxis, 37*, 172.

*Michopoulos, I., Douzenis, A., Gournellis, R., Christodoulou, C., Kalkavoura, C., Michalopoulou, P.G., … & Lykouras, L. (2010). Major depression in elderly medical inpatients in Greece, prevalence and identification. *Aging Clinical and Experimental Research, 22*, 148-151.

Mitchell, A. J., Meader, N., & Symonds, P. (2010). Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: A meta-analysis. *Journal of Affective Disorders, 126*(3), 335-348. doi:10.1016/j.jad.2010.01.067

National Collaborating Centre for Mental Health (UK). (2010). *Depression in Adults with a Chronic Physical Health Problem: Treatment and Management*. British Psychological Society (UK). http://www.ncbi.nlm.nih.gov/books/NBK82916/

NCSS. (2017). One ROC curve and cutoff analysis, Chapter 546. Retrieved September 20, 2021, from https://www.ncss.com/software/ncss/roc-curves-ncss/

Negeri, Z. F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., Thombs, B. D. . . . DEPRESsion Screening Data (DEPRESSD) PHQ Group. (2021). Accuracy of the Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: an updated systematic review and individual participant data meta-analysis. *BMJ*, 375:n2183. doi: https://doi.org/10.1136/bmj.n2183

Neupane, D., Levis, B., Bhandari, P. M., Thombs, B. D., Benedetti, A., & DEPRESsion Screening Data (DEPRESSD) Collaboration. (2021). Selective cutoff reporting in studies of the accuracy of the PHQ-9 and EPDS depressions screening tools: comparison of results based on published cutoffs versus all cutoffs using individual participant data meta-analysis. *International Journal of Epidemiology*, 30(3), e1873. doi: 10.1002/mpr.1873.

Nosen, E., & Woody, S. R. (2008). Chapter 8: Diagnostic assessment in research. In *Handbook of Research Methods in Abnormal and Clinical Psychology* (ed. D. McKay), pp. 109-124. Sage: Thousand Oaks.

*O'Rourke, S., MacHale, S., Signorini, D., Dennis, M. (1998). Detecting psychiatric morbidity after stroke: comparison of the GHQ and the HAD Scale. *Stroke*, *29*, 980-985.

*Öztürk, A., Deveci, E., Bağcıoğlu, E., Atalay, F. & Serdar, Z. (2013). Anxiety, depression, social phobia, and quality of life in Turkish patients with acne and their relationships with the severity of acne. *Turkish Journal of Medical Sciences, 43* (4), 660-666.

*Patel, D., Sharpe, L., Thewes, B., Rickard, J., Schnieden, V., & Lewis, C. (2010). Feasibility of using risk factors to screen for psychological disorder during routine breast care nurse consultations. *Cancer Nursing, 33*(1), 19–27.

*Patel, D., Sharpe, L.A., Thewes, B., Bell, M.L., & Clarke, S.J. (2011). Using the Distress Thermometer and Hospital Anxiety and Depression Scale to screen for psychosocial morbidity in patients diagnosed with colorectal cancer. *Journal of Affective Disorders, 131*(1-3), 412-6.

*Patten, S. B., Burton, J. M., Fiest, K. M., Wiebe, S., Bulloch, A. G., Koch, M., … & Jetté, N. (2015). Validity of four screening scales for major depression in MS. *Multiple Sclerosis, 21*(8), 1064–1071.

*Pedroso, V. S., Vieira, É. L., Brunoni, A. R., Lauterbach, E. C., Teixeira, A. L. (2016). Psychopathological evaluation and use of the Hospital Anxiety and Depression Scale in a sample of Brazilian patients with post-stroke depression. *Archives of Clinical Psychiatry (São Paulo), 43*, 147-50.

*Phan, T., Carter, O., Adams, C., Waterer, G., Chung, L. P., Hawkins, M., … & Strobel, N. (2016). Discriminant validity of the Hospital Anxiety and Depression Scale, Beck Depression Inventory (II) and Beck Anxiety Inventory to confirmed clinical diagnosis of depression and anxiety in patients with chronic obstructive pulmonary disease. *Chronic Respiratory Disease, 13*(3), 220–228.

*Pintor, L., Fuente, E.D., Peri, J.M., Pérez-Villa, F., & Roig, E. (2006). Evaluación psiquiátrica transversal en pacientes candidatos a un trasplante cardíaco. *Psiquiatría Biológica, 13*, 122-126.

*Prisnie, J. C., Fiest, K. M., Coutts, S. B., Patten, S. B., Atta, C. A., Blaikie, L., … & Jetté, N. (2016). Validating screening tools for depression in stroke and transient ischemic attack patients. *International Journal of Psychiatry in Medicine, 51*(3), 262–277.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA http://www.rstudio.com/.

*Reme, S. E., Lie, S. A., & Eriksen, H. R. (2014). Are 2 questions enough to screen for depression and anxiety in patients with chronic low back pain?. *Spine, 39*(7), E455–E462.

Rice, D. B., & Thombs, B. D. (2016). Risk of Bias from Inclusion of Currently Diagnosed or Treated Patients in Studies of Depression Screening Tool Accuracy: A Cross-Sectional Analysis of Recently Published Primary Studies and Meta-Analyses. *Plos One, 11*(2). doi:10.1371/journal.pone.0150067

Riley, R. D., Dodd, S. R., Craig, J. V., Thompson, J. R., & Williamson, P. R. (2008). Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Statistics in Medicine, 27*(29), 6111-6136. doi:10.1002/sim.3441

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ, 340*. doi:10.1136/bmj.c221

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., . . . Towle, L. H. (1988). The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives Of General Psychiatry, 45*(12), 1069-1077.

*Rooney, A. G., McNamara, S., Mackinnon, M., Fraser, M., Rampling, R., Carson, A., & Grant, R. (2013). Screening for major depressive disorder in adults with cerebral glioma: an initial validation of 3 self-report instruments. *Neuro-Oncology, 15*(1), 122-129.

*Ryan, D. A., Gallagher, P., Wright, S., Cassidy, E. M. (2012). Sensitivity and specificity of the Distress Thermometer and a two-item depression screen (Patient Health Questionnaire-2) with a 'help' question for psychological distress and psychiatric morbidity in patients with advanced cancer. *Psycho-Oncology, 21*(12), 1275.

*Sanchez-Gistau, V., Sugranyes, G., Baillés, E., Carreño, M., Donaire, A., Bargalló, N., & Pintor, L. (2012). Is major depressive disorder specifically associated with mesial temporal sclerosis?. *Epilepsia, 53*(2), 386–392.

*Sánchez, R., Peri, J. M., Baillés, E., Bastidas, A., Pérez-Villa, F., Bulbena, A., & Pintor, L. (2012). Evaluación de psicopatología, afrontamiento y apoyo familiar en el cumplimiento de pautas médicas en los 12 meses posteriores a un trasplante cardiaco. *Psiquiatría Biológica, 19*, 1-5.

*Sánchez, R., Baillés, E., Peri, J.M., Bastidas, A., Pérez-Villa, F., Bulbena, A., & Pintor, L. (2014). Cross-sectional psychosocial evaluation of heart transplantation candidates. *General Hospital Psychiatry, 36*(6), 680-5.

*Saracino, R.M., Weinberger, M.I., Roth, A.J., Hurria, A., & Nelson, C.J. (2017). Assessing depression in a geriatric cancer population. *Psycho-Oncology, 26*, 1484 - 1490.

Schatzberg, A. F. (2019). Scientific Issues Relevant to Improving the Diagnosis, Risk Assessment, and Treatment of Major Depression. *American Journal of Psychiatry, 176*(5), 342-347. doi:10.1176/appi.ajp.2019.19030273

*Schellekens, M.P., van den Hurk, D.G., Prins, J.B., Molema, J., van der Drift, M.A., & Speckens, A.E. (2016). The suitability of the Hospital Anxiety and Depression Scale, Distress Thermometer and other instruments to screen for psychiatric disorders in both lung cancer patients and their partners. *Journal of Affective Disorders, 203*, 176-183 .

*Schwarzbold, M.L., Diaz, A.P., Nunes, J.C., Sousa, D.S., Hohl, A., Guarnieri, R., … & Walz, R. (2014). Validity and screening properties of three depression rating scales in a prospective sample of patients with severe traumatic brain injury. *Revista brasileira de psiquiatria, 36*, 206-12 .

*Senturk, V., Stewart, R.J., & Sağduyu, A. (2007). Screening for mental disorders in leprosy patients: comparing the internal consistency and screening properties of HADS and GHQ-12. *Leprosy Review, 78*(3), 231-42 .

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., . . . Dunbar, G. C. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry, 12*(5), 232-241. doi:10.1016/S0924-9338(97)83297-X

*Sia, A.D., Williams, L.J., Pasco, J.A., Jacka, F.N., Brennan-Olsen, S.L., & Veerman, J.L. (2018). The Population Mean Mood Predicts The Prevalence of Depression in an Australian Context. *Australian & New Zealand Journal of Psychiatry, 52,* 461 - 472.

*Simard, S., & Savard, J. (2015). Screening and comorbidity of clinical levels of fear of cancer recurrence. *Journal of Cancer Survivorship, 9*, 481-491.

*Singer, S., Danker, H., Dietz, A., Hornemann, B., Koscielny, S., Oeken, J., … & Krauss, O. (2008). Screening for mental disorders in laryngeal cancer patients: a comparison of 6 methods. *Psycho-Oncology, 17.*

*Singer, S., Kuhnt, S., Götze, H., Hauss, J., Hinz, A., Liebmann, A., Krauss, O., Lehmann, A., & Schwarz, R. (2009). Hospital anxiety and depression scale cutoff scores for cancer patients in acute care. *British Journal of Cancer, 100*(6), 908–912.

Siu, A. L., & US Preventive Services Task Force (USPSTF). (2016). Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA*, *315*(4), 380–387. https://doi.org/10.1001/jama.2015.18392

*Soyseth, T. S., Lund, M. B., Bjortuft, O., Heldal, A., Søyseth, V., Dew, M. A., … & Malt, U. F. (2016). Psychiatric disorders and psychological distress in patients undergoing evaluation for lung transplantation: a national cohort study. *General Hospital Psychiatry*, *42*, 67–73.

*Stafford, L., Berk, M., & Jackson, H. J. (2007). Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. *General Hospital Psychiatry*, *29*(5), 417–424.

*Stafford, L., Judd, F., Gibson, P., Komiti, A., Quinn, M., & Mann, G. B. (2014). Comparison of the hospital anxiety and depression scale and the center for epidemiological studies depression scale for detecting depression in women with breast or gynecologic cancer. *General Hospital Psychiatry, 36*(1), 74–80.

*Stone, J., Townend, E., Kwan, J., Haga, K., Dennis, M. S., & Sharpe, M. (2004). Personality change after stroke: Some preliminary observations. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(12), 1708–1713.

*Sultan, S., Luminet, O., & Hartemann, A. (2010). Cognitive and anxiety symptoms in screening for clinical depression in diabetes: a systematic examination of diagnostic performances of the HADS and BDI-SF. *Journal of Affective Disorders*, *123*(1-3), 332–336.

Thombs, B. D., Arthurs, E., El-Baalbaki, G., Meijer, A., Ziegelstein, R. C., & Steele, R. J. (2011). Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review. *BMJ, 343*. doi:10.1136/bmj.d4825

Thombs, B. D., Benedetti, A., Kloda, L. A., Levis, B., Azar, M., Riehm, K. E., . . . Tonelli, M. (2016). Diagnostic accuracy of the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. *BMJ Open, 6*(4). doi:10.1136/bmjopen-2016-011913

Thombs, B. D., & Rice, D. B. (2016). Sample sizes and precision of estimates of sensitivity and specificity from primary studies on the diagnostic accuracy of depression screening tools: a survey of recently published studies. *International Journal of Methods in Psychiatric Research, 25*(2), 145-152. doi:10.1002/mpr.1504

*Tiringer, I., Simon, A., Herrfurth, D., Suri, I., Szalai, K., & Veress, A. (2008). Occurrence of anxiety and depression disorders after acute cardiac events during hospital rehabilitation. Application of the Hospital Anxiety and Depression Scale as a screening instrument. *Psychiatria Hungarica, 23*(6), 430–443.

*Tung, K. Y., Cheng, K. S., Lee, W. K., Kwong, P. K., Chan, K. W., Law, A. C., & Lo, W. T. (2015). Psychiatric Morbidity in Chinese Adults with Type 1 Diabetes in Hong Kong. *East Asian Archives of Psychiatry, 25*(3), 128–136.

*Turner, A., Hambridge, J., White, J., Carter, G., Clover, K., Nelson, L., & Hackett, M. (2012). Depression screening in stroke: a comparison of alternative measures with the structured diagnostic interview for the diagnostic and statistical manual of mental disorders, fourth edition (major depressive episode) as criterion standard. *Stroke, 43*(4), 1000–1005.

United Nations Development Programme. (2020). Human Development Reports. http://hdr.undp.org/en/content/human-development-index-hdi. Accessed September 20, 2021.

van der Leeden R, Busing FMTA, Meijer E. (1997). *Bootstrap methods for two-level models. Technical report PRM 97-04.* Leiden University, Department of Psychology: Leiden, The Netherlands.

van der Leeden R, Meijer E, Busing FMTA (2008). Chapter 11: Resampling multilevel models. In *Handbook of Multilevel Analysis* (ed. J. Leeuw, E. Meijer), pp. 401-433. Springer: New York.

Vodermaier, A., & Millman, R. D. (2011). Accuracy of the Hospital Anxiety and Depression Scale as a screening tool in cancer patients: a systematic review and meta-analysis. *Supportive Care in Cancer, 19*(12), 1899-1908. doi:10.1007/s00520-011-1251-4

*Walker, J., Postma, K., McHugh, G. S., Rush, R., Coyle, B., Strong, V., & Sharpe, M. (2007). Performance of the Hospital Anxiety and Depression Scale as a screening tool for major depressive disorder in cancer patients. *Journal of Psychosomatic Research, 63*(1), 83–91.

*Walterfang, M. A., O'Donovan, J., Fahey, M. C., & Velakoulis, D. (2007). The neuropsychiatry of adrenomyeloneuropathy. *CNS Spectrums, 12*(9), 696–701.

*Wong, L. Y., Yiu, R. L., Chiu, C. K., Lee, W. K., Lee, Y. L., Kwong, P. K., & Lo, W. T. (2015). Prevalence of Psychiatric Morbidity in Chinese Subjects with Knee Osteoarthritis in a Hong Kong Orthopaedic Clinic. *East Asian Archives of Psychiatry, 25*(4), 150–158.

Walker, E., & Nowacki, A. S. (2011). Understanding Equivalence and Noninferiority Testing. *Journal of General Internal Medicine, 26*(2), 192-196. doi:10.1007/s11606-010-1513-8

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . . QUADAS-2 Group. (2011). QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine, 155*(8), 529-U104. doi:10.7326/0003-4819-155-8-201110180-00009

World Health Organization (WHO). (1992). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. World Health Organization.

World Health Organization. (1994). *Schedules for clinical assessment in neuropsychiatry: version 2*. World Health Organization.

Wu, Y., Levis, B., Ioannidis, J. P. A., Benedetti, A., Thombs, B. D., & DEPRESsion Screening Data (DEPRESSD) Collaboration. (2020). Probability of Major Depression Classification Based on the SCID, CIDI, and MINI Diagnostic Interviews: A Synthesis of Three Individual Participant Data Meta-Analyses. *Psychotherapy and Psychosomatics, 90*(1), 28-40. doi:10.1159/000509283

Wu, Y., Levis, B., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., . . . Thombs, B. D. (2020). Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. *Psychological Medicine, 50*(8), 1368-1380. doi:10.1017/S0033291719001314

Wu, Y., Levis, B., Sun, Y., He, C., Krishnan, A., Neupane, D., . . . DEPRESsion Screening Data (DEPRESSD) HADS Group. (2021). Accuracy of the Hospital Anxiety and Depression Scale Depression subscale (HADS-D) to screen for major depression: systematic review and individual participant data meta-analysis. *BMJ, 373*. doi:10.1136/bmj.n972

Wu, Y., Levis, B., Sun, Y., Krishnan, A., He, C., Riehm, K. E., . . . Thombs, B. D. (2020). Probability of major depression diagnostic classification based on the SCID, CIDI and MINI diagnostic interviews controlling for Hospital Anxiety and Depression Scale - Depression subscale scores: An individual participant data meta-analysis of 73 primary studies. *Journal of Psychosomatic Research, 129*. doi:10.1016/j.jpsychores.2019.109892

*Yamashita, A., Noguchi, H., Hamazaki, K., Sato, Y., Narisawa, T., Kawashima, Y., … & Matsuoka, Y. J. (2017). Serum polyunsaturated fatty acids and risk of psychiatric disorder after acute coronary syndrome: A prospective cohort study. *Journal of Affective Disorders*, *218*, 306–312.

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67*(6), 361–370. doi: 10.1111/j.1600-0447.1983.tb09716.x

*Studies that included in the IPDMA

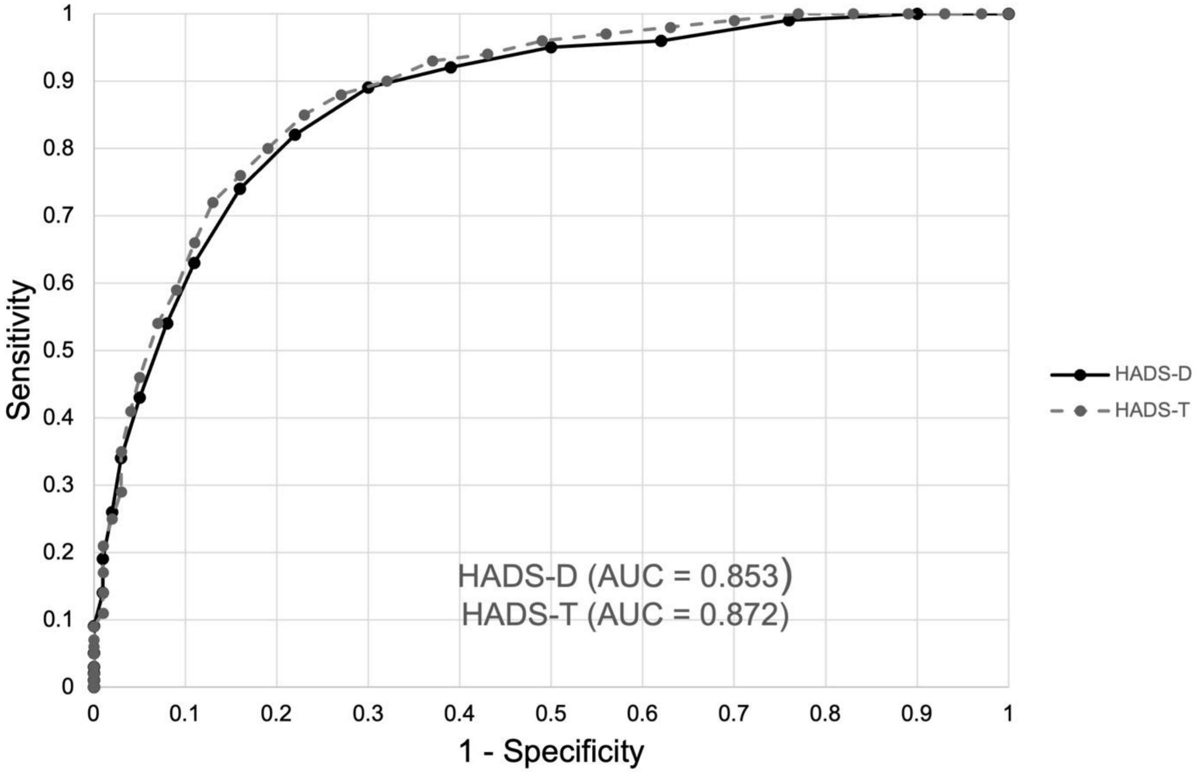
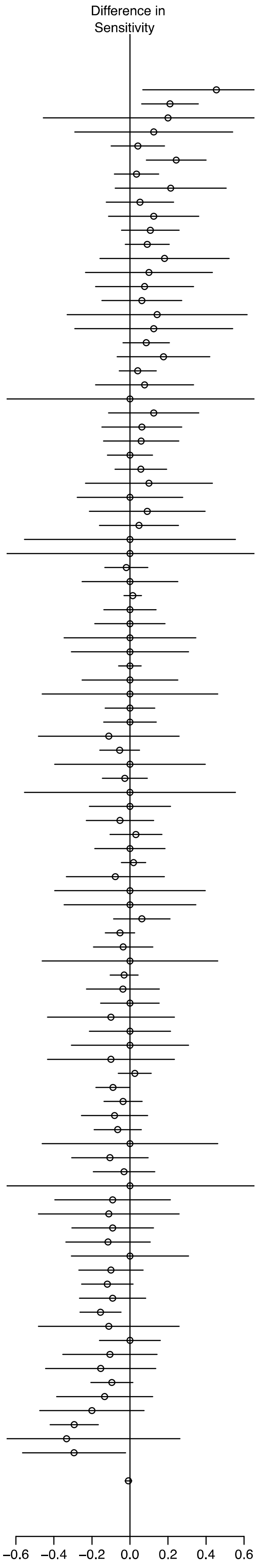


Fig 1. ROC curve for HADS-D and HADS-T across all studies.

Study	MDD/Total N (Weighted)	Difference in Sensitivity (95% CI)	Difference in Specificity (95% CI)
Pedroso, 2016 [88]	9 / 48	0.45 (0.07 , 0.84)	0.12 (0.00 , 0.24)
Kang, 2013 [81]	36 / 423	0.21 (0.06 , 0.36)	0.16 (0.12 , 0.20)
Sanchez, 2012 [41]	3 / 22	0.20 (−0.46 , 0.86)	0.14 (−0.06 , 0.34)
Senturk, 2007 [65]	6 / 57	0.12 (−0.29 , 0.54)	0.21 (0.09 , 0.33)
Huey, 2018 [19]	22 / 236	0.04 (−0.10 , 0.18)	0.21 (0.15 , 0.27)
Sanchez−Gistau, 2012 [40]	35 / 296	0.24 (0.09 , 0.40)	0.00 (−0.05 , 0.05)
Michopoulos, 2010 [32]	27 / 193	0.03 (−0.08 , 0.15)	0.20 (0.13 , 0.27)
De Souza, 2009 [9]	12 / 50	0.21 (−0.08 , 0.51)	0.00 (−0.12 , 0.12)
Akechi, 2006 [1]	17 / 223	0.05 (−0.12 , 0.23)	0.16 (0.10 , 0.22)
Cukor, 2008 [7]	14 / 70	0.12 (−0.11 , 0.36)	0.07 (−0.01 , 0.15)
Matsuoka, 2009 [86]	26 / 153	0.11 (−0.04 , 0.26)	0.07 (0.02 , 0.12)
Beck, 2016 [67]	53 / 313	0.09 (−0.02 , 0.21)	0.08 (0.04 , 0.12)
Honarmand, 2009 [18]	9 / 140	0.18 (−0.16 , 0.52)	−0.02 (−0.07 , 0.04)
Ferentinos, 2011 [11]	8 / 36	0.10 (−0.23 , 0.43)	0.07 (−0.06 , 0.20)
Jang, 2012 [80]	11 / 309	0.08 (−0.18 , 0.33)	0.08 (0.05 , 0.11)
Schwarzbold, 2014 [45]	14 / 44	0.06 (−0.15 , 0.27)	0.09 (−0.04 , 0.23)
Yamashita, 2017 [96]	5 / 98	0.14 (−0.33 , 0.62)	0.01 (−0.04 , 0.06)
Saracino, 2017 [43]	6 / 188	0.12 (−0.29 , 0.54)	0.02 (−0.02 , 0.06)
Wong, 2015 [56]	33 / 114	0.09 (−0.04 , 0.21)	0.06 (−0.01 , 0.13)
Rooney, 2013 [38]	15 / 133	0.18 (−0.07 , 0.42)	−0.03 (−0.09 , 0.02)
Chen, 2010 [71]	47 / 195	0.04 (−0.06 , 0.14)	0.09 (0.03 , 0.14)
Fischer, 2014 [13]	11 / 194	0.08 (−0.18 , 0.33)	0.05 (0.00 , 0.10)
Cheung, 2011 [72]	1 / 55	0.00 (−0.92 , 0.92)	0.12 (0.01 , 0.24)
Gagnon, 2005 [14]	14 / 108	0.12 (−0.11 , 0.36)	0.00 (−0.06 , 0.06)
da Rocha e Silva, 2013 [8]	14 / 47	0.06 (−0.15 , 0.27)	0.06 (−0.05 , 0.17)
Gould, 2011 [16]	15 / 189	0.06 (−0.14 , 0.26)	0.06 (0.01 , 0.11)
Juliao, 2013 [20]	31 / 75	0.00 (−0.12 , 0.12)	0.11 (−0.02 , 0.23)
Tung, 2015 [51]	33 / 136	0.06 (−0.08 , 0.19)	0.05 (−0.01 , 0.11)
Sanchez, 2014 [42]	8 / 120	0.10 (−0.23 , 0.43)	−0.01 (−0.08 , 0.06)
Loosman, 2010 [84]	8 / 28	0.00 (−0.28 , 0.28)	0.09 (−0.12 , 0.31)
O'Rourke, 1998 [33]	9 / 56	0.09 (−0.21 , 0.39)	0.00 (−0.10 , 0.10)
Patten, 2015 [35]	19 / 41	0.05 (−0.16 , 0.26)	0.04 (−0.17 , 0.26)
Kugaya, 2000 [23]	3 / 81	0.00 (−0.55 , 0.55)	0.09 (−0.01 , 0.18)
Braeken, 2010 [5]	1 / 12	0.00 (−0.92 , 0.92)	0.08 (−0.18 , 0.33)
Dorow, 2017 [10]	50 / 1143	−0.02 (−0.13 , 0.09)	0.08 (0.06 , 0.10)
Beraldi, 2014 [3]	9 / 117	0.00 (−0.25 , 0.25)	0.06 (−0.01 , 0.14)
Butnoriene, 2014 [70]	201 / 1115	0.01 (−0.03 , 0.06)	0.04 (0.02 , 0.07)
Fabregas, 2014 [78]	33 / 105	0.00 (−0.14 , 0.14)	0.05 (−0.02 , 0.13)
Douven, 2016 [76]	13 / 247	0.00 (−0.18 , 0.18)	0.05 (0.01 , 0.08)
Phan, 2016 [89]	6 / 47	0.00 (−0.35 , 0.35)	0.05 (−0.06 , 0.16)
Jackson, Unpublished	7 / 52	0.00 (−0.31 , 0.31)	0.04 (−0.10 , 0.19)
McFarlane, 2009 [87]	130 / 859	0.00 (−0.06 , 0.06)	0.04 (0.01 , 0.06)
Soyseth, 2016 [91]	9 / 94	0.00 (−0.25 , 0.25)	0.03 (−0.03 , 0.10)
Stone, 2004 [50]	4 / 35	0.00 (−0.46 , 0.46)	0.03 (−0.10 , 0.16)
Harter, 2006 [61]	28 / 512	0.00 (−0.13 , 0.13)	0.03 (−0.00 , 0.06)
Hahn, 2006 [60]	18 / 205	0.00 (−0.14 , 0.14)	0.03 (−0.03 , 0.08)
Can, 2018 [6]	7 / 141	−0.11 (−0.48 , 0.26)	0.13 (0.07 , 0.19)
de Oliveira, 2014 [75]	35 / 126	−0.05 (−0.16 , 0.05)	0.08 (−0.01 , 0.16)
Patel, 2010 [63]	5 / 52	0.00 (−0.40 , 0.40)	0.02 (−0.07 , 0.11)
Gandy, 2012 [79]	35 / 147	−0.03 (−0.15 , 0.09)	0.04 (−0.02 , 0.11)
Drabe, 2008 [77]	3 / 62	0.00 (−0.55 , 0.55)	0.02 (−0.04 , 0.07)
Grassi, 2009 [59]	11 / 301	0.00 (−0.21 , 0.21)	0.01 (−0.03 , 0.05)
Reme, 2014 [90]	17 / 537	−0.05 (−0.23 , 0.12)	0.06 (0.03 , 0.08)
Walker, 2007 [54]	30 / 361	0.03 (−0.11 , 0.17)	−0.03 (−0.06 , 0.01)
Pintor, 2006 [36]	13 / 73	0.00 (−0.18 , 0.18)	0.00 (−0.09 , 0.09)
Amoozegar, 2017 [2]	51 / 101	0.02 (−0.04 , 0.08)	−0.02 (−0.12 , 0.08)
Lees, 2013 [83]	11 / 65	−0.08 (−0.33 , 0.18)	0.07 (−0.03 , 0.17)
Lee, 2016 [25]	5 / 106	0.00 (−0.40 , 0.40)	−0.01 (−0.06 , 0.04)
Lee, 2017 [26]	6 / 143	0.00 (−0.35 , 0.35)	−0.01 (−0.06 , 0.03)
Law, 2014 [82]	30 / 100	0.06 (−0.09 , 0.21)	−0.08 (−0.19 , 0.02)
Hartung, 2017 [62]	87 / 1393	−0.05 (−0.13 , 0.03)	0.03 (0.01 , 0.05)
Marrie, 2018 [30]	26 / 252	−0.04 (−0.19 , 0.12)	0.01 (−0.04 , 0.06)
Keller, 2004 [21]	4 / 76	0.00 (−0.46 , 0.46)	−0.03 (−0.09 , 0.04)
Lowe, 2002 [29]	63 / 490	−0.03 (−0.10 , 0.04)	0.00 (−0.03 , 0.03)
Lambert, 2015 [24]	25 / 164	−0.04 (−0.23 , 0.15)	0.00 (−0.05 , 0.05)
Love, 2004 [28]	16 / 227	0.00 (−0.15 , 0.15)	−0.04 (−0.09 , 0.01)
Singer, 2008 [48]	8 / 141	−0.10 (−0.43 , 0.23)	0.05 (0.00 , 0.10)
Costa−Requena, 2013 [58]	11 / 192	0.00 (−0.21 , 0.21)	−0.05 (−0.10 , 0.00)
Simard, 2015 [47]	7 / 60	0.00 (−0.31 , 0.31)	−0.05 (−0.13 , 0.02)
Ryan, 2012 [39]	8 / 203	−0.10 (−0.43 , 0.23)	0.04 (−0.01 , 0.09)
Al−Asmi, 2011 [57]	37 / 140	0.03 (−0.06 , 0.11)	−0.09 (−0.16 , −0.02)
Singer, 2009 [49]	54 / 576	−0.09 (−0.18 , 0.00)	0.03 (−0.00 , 0.06)
Sia, 2018 [46]	53 / 789	−0.04 (−0.14 , 0.06)	−0.03 (−0.05 , −0.01)
Stafford, 2007 [92]	35 / 193	−0.08 (−0.25 , 0.09)	0.01 (−0.04 , 0.05)
Sultan, 2009 [94]	29 / 282	−0.06 (−0.19 , 0.06)	−0.01 (−0.06 , 0.03)
Meyer, 2008 [31]	4 / 102	0.00 (−0.46 , 0.46)	−0.08 (−0.16 , 0.00)
Hitchon, 2019 [17]	17 / 149	−0.11 (−0.31 , 0.10)	0.01 (−0.04 , 0.07)
Fiest, 2014 [12]	30 / 179	−0.03 (−0.19 , 0.13)	−0.06 (−0.12 , −0.00)
Walterfang, 2007 [55]	1 / 10	0.00 (−0.92 , 0.92)	−0.09 (−0.39 , 0.21)
Tiringer, 2008 [95]	9 / 143	−0.09 (−0.39 , 0.21)	−0.01 (−0.06 , 0.04)
Patel, 2011 [64]	7 / 92	−0.11 (−0.48 , 0.26)	0.01 (−0.06 , 0.09)
Kjaergaard, 2014 [22]	20 / 357	−0.09 (−0.31 , 0.12)	−0.01 (−0.03 , 0.01)
Bayon−Perez, 2016 [66]	24 / 113	−0.12 (−0.34 , 0.11)	0.01 (−0.05 , 0.07)
Golden, 2006 [15]	7 / 85	0.00 (−0.31 , 0.31)	−0.11 (−0.21 , −0.01)
Love, 2002 [27]	28 / 302	−0.10 (−0.27 , 0.07)	−0.02 (−0.06 , 0.02)
Bunevicius, 2007 [68]	40 / 494	−0.12 (−0.25 , 0.02)	−0.00 (−0.03 , 0.03)
Bernstein, 2018 [4]	20 / 245	−0.09 (−0.26 , 0.08)	−0.04 (−0.08 , −0.00)
De la Torre, 2016 [74]	69 / 256	−0.15 (−0.26 , −0.05)	0.02 (−0.03 , 0.07)
Ozturk, 2013 [34]	7 / 45	−0.11 (−0.48 , 0.26)	−0.02 (−0.13 , 0.08)
Massardo, 2015 [85]	28 / 128	0.00 (−0.16 , 0.16)	−0.14 (−0.22 , −0.06)
Stafford, 2014 [93]	17 / 100	−0.11 (−0.35 , 0.14)	−0.04 (−0.12 , 0.05)
Prisnie, 2016 [37]	11 / 114	−0.15 (−0.44 , 0.14)	0.01 (−0.07 , 0.09)
Sanchez, Unpublished	40 / 394	−0.10 (−0.21 , 0.02)	−0.05 (−0.09 , −0.01)
Turner, 2012 [52]	13 / 72	−0.13 (−0.39 , 0.12)	−0.02 (−0.10 , 0.07)
Schellekens, 2016 [44]	13 / 151	−0.20 (−0.47 , 0.07)	0.03 (−0.02 , 0.08)
Bunevicius, 2012 [69]	56 / 517	−0.29 (−0.42 , −0.17)	−0.02 (−0.04 , 0.01)
Turner, Unpublished [53]	4 / 52	−0.33 (−0.93 , 0.26)	−0.04 (−0.12 , 0.04)
Consoli, 2006 [73]	15 / 93	−0.29 (−0.57 , −0.02)	−0.25 (−0.36 , −0.14)
Pooled − Random Effects	2285 / 20700	−0.01 (−0.03 , 0.01)	0.02 (0.01 , 0.03)



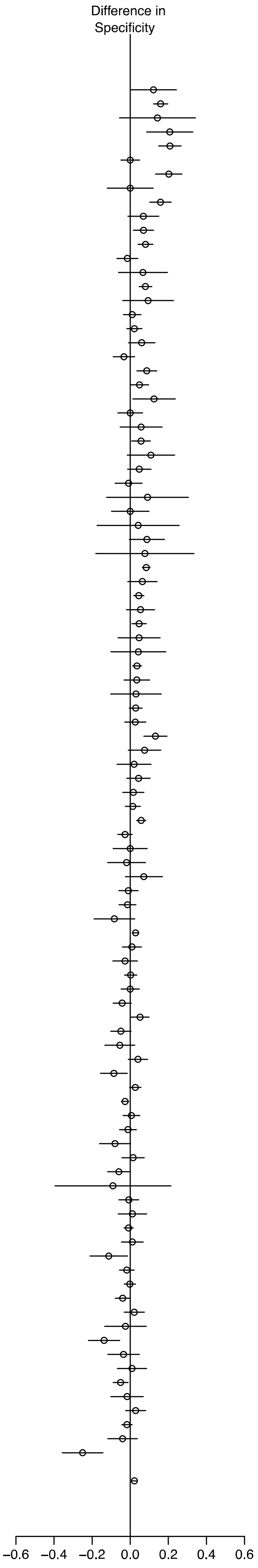


Fig 2. Forest plots of the difference in sensitivity and specificity estimates at the optimal cutoff (HADS-D: ≥7; HADS-T: ≥15) between HADS-D and HADS-T across all studies[a] (N Studies = 98[b]; N Participants = 20,700; N major depression = 2,285)[c]

[a] τ^2 for the difference of sensitivity and specificity were both <0.001.
[b] References for all included studies are marked with an asterisk in the reference list. The reference numbers refer to Supplementary Material References.
[c] The studies were sorted by the sum of difference in sensitivity and difference in specificity in descending order.

Table 1. Participant data by subgroups[a]

Participant Subgroup	N Studies	N Participants	N (%) Major Depression
All participants	98	20,700	2,285 (11)
Participants not currently diagnosed with a mental disorder or receiving treatment for a mental health problem	38	6,995	495 (7)
Age <60	92	11,795	1,452 (12)
Age ≥60	92	8,741	779 (9)
Women	96	11,111	1,342 (12)
Men	89	9,494	911 (10)
Very high country human development index	90	20,088	2,130 (11)
High country human development index	8	612	155 (25)
Participants diagnosed with cancer[b]	27	5,767	433 (8)
Inpatient specialty care	38	8,827	1,047 (12)
Outpatient specialty care	54	9,547	1,072 (11)
Non-medical	7	1,908	116 (6)
Inpatient/outpatient mixed	3	418	50 (12)

[a] Some variables were coded at the study level, while others were coded at the participant level. Thus, number of studies does not always add up to the total number.

[b] The statistics here were from individual-level variable of cancer diagnosis, slight different from what we used in the subgroup analysis which based on the study-level care setting variable.

Table 2. Comparison of sensitivity and specificity estimates between HADS-D and HADS-T for pairs of optimal cutoffs and cutoffs close to the optimal cutoffs across all studies

HADS-D[a]					HADS-T					HADS-T – HADS-D			
Cutoff	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Cutoff**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	0.90	(0.87, 0.92)	0.61	(0.58, 0.64)	11	0.91	(0.89, 0.93)	0.63	(0.60, 0.66)	0.01	(-0.01, 0.04)	0.02	(-0.00, 0.04)
6	0.86	(0.82, 0.88)	0.70	(0.67, 0.73)	13	0.86	(0.83, 0.88)	0.73	(0.70, 0.75)	0.00	(-0.03, 0.03)	0.03	(0.01, 0.05)
7[b]	0.79	(0.75, 0.83)	0.78	(0.75, 0.80)	15[c]	0.79	(0.76, 0.82)	0.81	(0.78, 0.83)	0.00	(-0.05, 0.02)	0.03	(0.01, 0.04)
8	0.70	(0.66, 0.74)	0.84	(0.82, 0.86)	17	0.70	(0.66, 0.74)	0.87	(0.85, 0.89)	0.00	(-0.05, 0.04)	0.03	(0.01, 0.04)
9	0.60	(0.55, 0.64)	0.89	(0.87, 0.91)	19	0.58	(0.54, 0.61)	0.91	(0.9, 0.93)	-0.02	(-0.07, 0.02)	0.02	(0.01, 0.03)
10	0.50	(0.45, 0.54)	0.92	(0.91, 0.94)	21	0.45	(0.41, 0.49)	0.95	(0.94, 0.95)	-0.05	(-0.10, -0.01)	0.03	(0.01, 0.03)
11	0.39	(0.35, 0.43)	0.95	(0.94, 0.96)	23	0.34	(0.31, 0.37)	0.97	(0.96, 0.97)	-0.05	(-0.10, -0.01)	0.02	(0.01, 0.03)

[a] N Studies = 98; N Participants = 20,700; N major depression = 2,285
[b] The cutoff minimizes the values of the distance to the top-left corner of the ROC curves for HADS-D.
[c] The cutoff minimizes the values of the distance to the top-left corner of the ROC curves for HADS-T.

CI: confidence interval

Table 3. Comparison of sensitivity and specificity estimates between HADS-D and HADS-T for pairs of optimal cutoffs and cutoffs close to the optimal cutoffs across all studies via individual-level model

HADS-D[a]	**HADS-T**	**HADS-T – HADS-D**	
Cutoff	**Cutoff**	**Sensitivity**	**Specificity**
5	11	0.02 (-0.00, 0.03)	0.01 (-0.00, 0.03)
6	13	0.01 (-0.01, 0.03)	0.03 (0.01, 0.04)
7[b]	15[c]	0.00 (-0.02, 0.03)	0.02 (0.01, 0.04)
8	17	0.00 (-0.03, 0.03)	0.03 (0.02, 0.04)
9	19	-0.02 (-0.05, 0.01)	0.03 (0.02, 0.04)
10	21	-0.05 (-0.08, -0.02)	0.03 (0.02, 0.03)
11	23	-0.05 (-0.09, -0.02)	0.02 (0.02, 0.03)

[a] N Participants = 20,700; N major depression = 2,285
[b] The cutoff minimizes the values of the distance to the top-left corner of the ROC curves for HADS-D.
[c] The cutoff minimizes the values of the distance to the top-left corner of the ROC curves for HADS-T.

Table 4a. Comparison of sensitivity and specificity estimates between HADS-D and HADS-T for pairs of optimal cutoffs and cutoffs close to the optimal cutoffs among participants recruited from inpatient care settings

HADS-D[a]					HADS-T					HADS-T – HADS-D			
Cutoff	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Cutoff**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	0.90	(0.87, 0.93)	0.55	(0.49, 0.60)	11	0.90	(0.87, 0.92)	0.62	(0.56, 0.68)	0.00	(-0.03, 0.03)	0.07	(0.04, 0.11)
6	0.86	(0.83, 0.89)	0.64	(0.58, 0.69)	13	0.85	(0.81, 0.88)	0.72	(0.67, 0.77)	-0.01	(-0.07, 0.02)	0.08	(0.06, 0.12)
7[b]	0.80	(0.75, 0.83)	0.73	(0.68, 0.78)	15[cd]	0.79	(0.74, 0.82)	0.81	(0.76, 0.85)	-0.01	(-0.08, 0.02)	0.08	(0.05, 0.11)
8	0.73	(0.68, 0.78)	0.80	(0.76, 0.84)	17	0.69	(0.64, 0.74)	0.87	(0.83, 0.90)	-0.04	(-0.11, 0.03)	0.07	(0.04, 0.09)
9	0.63	(0.58, 0.69)	0.86	(0.82, 0.89)	19	0.59	(0.54, 0.64)	0.91	(0.88, 0.93)	-0.04	(-0.14, 0.01)	0.05	(0.03, 0.07)
10	0.55	(0.49, 0.61)	0.90	(0.87, 0.93)	21	0.46	(0.41, 0.51)	0.95	(0.92, 0.96)	-0.09	(-0.19, -0.03)	0.05	(0.03, 0.06)
11	0.45	(0.39, 0.51)	0.93	(0.91, 0.95)	23	0.36	(0.32, 0.41)	0.97	(0.95, 0.98)	-0.09	(-0.18, -0.02)	0.04	(0.02, 0.05)

[a] N Studies = 38; N Participants = 8,827; N major depression = 1,047
[b] The cutoff minimizes the values of the distance to the top-left corner of the ROC curves for HADS-D.
[c] The cutoff minimizes the values of the distance to the top-left corner of the ROC curves for HADS-T.
[d] On this cutoff of HADS-T, the model convergence code was 0 when using the default optimizer in glmer, but there were meaningful CIs.

CI: confidence interval

Table 4b. Comparison of sensitivity and specificity estimates between HADS-D and HADS-T for pairs of optimal cutoffs and cutoffs close to the optimal cutoffs among participants recruited from outpatient care settings

HADS-D[a]					HADS-T					HADS-T – HADS-D			
Cutoff	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Cutoff**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
5	0.91	(0.87, 0.94)	0.63	(0.60, 0.67)	11	0.92	(0.89, 0.95)	0.62	(0.59, 0.66)	0.01	(-0.02, 0.04)	-0.01	(-0.03, 0.01)
6	0.87	(0.82, 0.91)	0.72	(0.69, 0.75)	13	0.88	(0.84, 0.91)	0.72	(0.69, 0.75)	0.01	(-0.02, 0.05)	0.00	(-0.01, 0.02)
7[b]	0.82	(0.75, 0.86)	0.79	(0.76, 0.81)	15[c]	0.81	(0.76, 0.84)	0.80	(0.77, 0.82)	-0.01	(-0.07, 0.04)	0.01	(-0.01, 0.03)
8	0.71	(0.65, 0.77)	0.85	(0.83, 0.87)	17	0.73	(0.67, 0.78)	0.86	(0.84, 0.88)	0.02	(-0.04, 0.07)	0.01	(-0.00, 0.03)
9	0.60	(0.54, 0.66)	0.90	(0.88, 0.91)	19	0.59	(0.53, 0.65)	0.91	(0.90, 0.92)	-0.01	(-0.08, 0.04)	0.01	(0.00, 0.03)
10	0.49	(0.43, 0.55)	0.93	(0.91, 0.94)	21	0.45	(0.39, 0.52)	0.94	(0.93, 0.95)	-0.04	(-0.11, 0.02)	0.01	(0.00, 0.03)
11	0.38	(0.32, 0.44)	0.95	(0.94, 0.96)	23	0.34	(0.29, 0.39)	0.96	(0.95, 0.97)	-0.04	(-0.10, 0.01)	0.01	(0.00, 0.02)

[a] N Studies = 54; N Participants = 9,547; N major depression = 1,072
[b] The cutoff minimizes the values of the distance to the top-left corner of the ROC curves for HADS-D.
[c] The cutoff minimizes the values of the distance to the top-left corner of the ROC curves for HADS-T.

CI: confidence interval