

Supplementary Online Content

Levis B, Sun Y, He C, et al; Depression Screening Data (DEPRESSD) PHQ Collaboration. Accuracy of the PHQ-2 alone and in combination with the PHQ-9 for screening to detect major depression: systematic review and meta-analysis. *JAMA*. doi:10.1001/jama.2020.6504

eMethods 1. Search Strategies

eMethods 2. Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) Coding Manual for Primary Studies Included in the Present Study

eFigure 1. Receiver operating characteristic (ROC) Plots Comparing Sensitivity and Specificity Estimates for Patient Health Questionnaire-2 (PHQ-2) Cutoffs 1-6 Among Semi-structured Diagnostic Interviews, Fully Structured Diagnostic Interviews, and the Mini International Neuropsychiatric Interview (MINI)

eFigure 2. Forest Plots of Sensitivity and Specificity Estimates for Cutoff 2 and 3 of the PHQ-2 for Each Reference Standard Category, Including Among Participants Verified to Not Currently be Diagnosed or Receiving Treatment for a Mental Health Problem as Well as Among Participant Subgroups Based on Age, Sex, Human Development Index and Care Setting

eFigure 3. Nomograms of Positive and Negative Predictive Value for Assumed Major Depression Prevalence of 5-25%, Based on Accuracy Estimates Among Studies With a Semi-structured Reference Standard and PHQ-9 Scores Available

eTable 1. Characteristics of Included Primary Studies as Well as Eligible Primary Studies Not Included in the Present Study

eTable 2. Numbers of Participants and Cases of Major Depression by Diagnostic Interview

eTable 3. Estimates of Heterogeneity at PHQ-2 Cutoff Score of 2 and 3

eTable 4. Comparison of PHQ-2 Sensitivity and Specificity Estimates Among Participants Verified to Not Currently be Diagnosed or Receiving Treatment for a Mental Health Problem Compared to All Participants as Well as Among Participant Subgroups Based on Age, Sex, Human Development Index, Care Setting, and Risk of Bias Factors, for Each Reference Standard Category

eTable 5. Sensitivity and Specificity Estimates for the PHQ-2 Alone, the PHQ-9 Alone, and for PHQ-2 \geq 2 Followed by PHQ-9 Among 44 Studies (N Participants = 10,627; N Major Depression = 1,361) That Used a Semi-structured Reference Standard and Had Both PHQ-2 and PHQ-9 Item Scores Available

eTable 6. Comparison of Sensitivity and Specificity for PHQ-2 \geq 2 in Combination With PHQ-9 \geq 5 to 15 Versus Sensitivity and Specificity for PHQ-9 \geq 5 to 15, Among Studies That Used a Semi-structured Diagnostic Interview as the Reference Standard

eTable 7. QUADAS-2 Ratings for Each Primary Study Included in the Present Study

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods 1. Search strategies

MEDLINE (OvidSP)

1. PHQ*.af.
2. patient health questionnaire*.af.
3. 1 or 2
4. Mass Screening/
5. Psychiatric Status Rating Scales/
6. "Predictive Value of Tests"/
7. "Reproducibility of Results"/
8. exp "Sensitivity and Specificity"/
9. Psychometrics/
10. Prevalence/
11. Reference Values/
- 12.. Reference Standards/
13. exp Diagnostic Errors/
14. Mental Disorders/di, pc [Diagnosis, Prevention & Control]
15. Mood Disorders/di, pc [Diagnosis, Prevention & Control]
16. Depressive Disorder/di, pc [Diagnosis, Prevention & Control]
17. Depressive Disorder, Major/di, pc [Diagnosis, Prevention & Control]
18. Depression, Postpartum/di, pc [Diagnosis, Prevention & Control]
19. Depression/di, pc [Diagnosis, Prevention & Control]
20. validation studies.pt.
21. comparative study.pt.
22. screen*.af.
23. prevalence.af.
24. predictive value*.af.
25. detect*.ti.
26. sensitiv*.ti.
27. valid*.ti.
28. revalid*.ti.
29. predict*.ti.
30. accur*.ti.
31. psychometric*.ti.
32. identif*.ti.
33. specificit*.ab.
34. cut?off*.ab.
35. cut* score*.ab.
36. cut?point*.ab.
37. threshold score*.ab.
38. reference standard*.ab.
39. reference test*.ab.
40. index test*.ab.
41. gold standard.ab.
42. or/4-41
43. 3 and 42
44. limit 43 to yr="2000-Current"

PsycINFO (OvidSP)

1. PHQ*.af.
2. patient health questionnaire*.af.
3. 1 or 2
4. Diagnosis/
5. Medical Diagnosis/
6. Psychodiagnosis/
7. Misdiagnosis/
8. Screening/
9. Health Screening/
10. Screening Tests/
11. Prediction/
12. Cutting Scores/
13. Psychometrics/
14. Test Validity/
15. screen*.af.
16. predictive value*.af.
17. detect*.ti.
18. sensitiv*.ti.
19. valid*.ti.
20. revalid*.ti.
21. accura*.ti.
22. psychometric*.ti.
23. specificit*.ab.
24. cut?off*.ab.
25. cut* score*.ab.
26. cut?point*.ab.
27. threshold score*.ab.
28. reference standard*.ab.
29. reference test*.ab.
30. index test*.ab.
31. gold standard.ab.
32. or/4-31
33. 3 and 32
38. Limit 33 to "2000 to current"

Web of Science (Web of Knowledge)

#1: TS=(PHQ* OR "Patient Health Questionnaire*")

#2: TS= (screen* OR prevalence OR "predictive value*" OR detect* OR sensitiv* OR valid* OR revalid* OR predict* OR accura* OR psychometric* OR identif* OR specificit* OR cutoff* OR "cut off*" OR "cut* score*" OR cutpoint* OR "cut point*" OR "threshold score*" OR "reference standard*" OR "reference test*" OR "index test*" OR "gold standard")

#1 AND #2

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2000-2018

eMethods 2. Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) coding manual for primary studies included in the present study

Domain 1: Participant Selection

1. **Signalling question 1 – Was a consecutive or random sample of patients enrolled?:** Code as “yes” if a consecutive or random sample of participants were recruited for the study and the percentage of eligible participants who participate is $\geq 75\%$. If the study indicates that consecutive or random participants were recruited, but does not give an indication of the total number of eligible participants and how many agreed to participate in the study, this should be rated “unclear”. If the percentage of eligible participants included in the study was between $\geq 50\%$ and $< 75\%$, then this should also be marked as “unclear”. If a very low rate of eligible participants ($< 50\%$) were included in the study, this should be coded “no.” In “Notes”, please provide the relevant numbers and percentages used to make a determination. If a convenience sample of participants was recruited for the study or if the study was a case-control design, code as “no”.
2. **Signalling question 2 – Was a case-control design avoided?:** Code as “yes” if the study did not employ a case-control design. Code as “no” if the study used a case-control design.
3. **Signalling question 3 – Did the study avoid inappropriate exclusions?:** Inappropriate exclusions refer to situations where an important part of the screening population was excluded from the study based on characteristics that could be related to screening results. Code as “yes” if the study does not inappropriately exclude participants. Code as “no” if the study inappropriately excludes participants.
4. **Overall risk of bias:** Rate as “low”, “high”, or “unclear” as described in QUADAS-2. Please indicate factors in decision in “Notes”. NOTE: if signalling question 1 was coded “Unclear” the overall risk of bias is either a) Unclear, in cases where the denominator is not specified, or the percentage cannot be calculated, or method of participant selection is unclear OR b) Low, in cases where the percentage can be calculated, and is between 50-75%. If signalling question 1 is a “no” and signalling questions 2 and 3 are both “yes” then the risk of bias is coded “Unclear”.
5. **Applicability concerns:** Code as “low” if study excluded participants who were already diagnosed or treated for depression or if the study included these patients, but they can be excluded using the individual patient data. Also code as “low” if the study did not exclude participants already diagnosed with depression and the overall percentage of these participants is low (e.g., $\leq 2.0\%$ of total participants), even if there is not a variable to exclude them. Code “unclear” if the study did not exclude participants already diagnosed or treated for depression and it is not known how many diagnosed and treated patients were included or if the percentage is moderate (e.g., $> 2.0\%$ but $\leq 5.0\%$). Code “high” if already diagnosed and treated patients are included and make up $> 5.0\%$ of the total sample and there is not a variable to exclude them. Please see aggregated study information sheet to code this.

Domain 2: Index Test

1. **Signalling question 1 - Were the index test results interpreted without the knowledge of the results of the reference standard?:** Code this item as “N/A” for all studies, as the index test is scored and does not require interpretation.
2. **Signalling question 2 - If a threshold was used, was it pre-specified?:** Code this item as “N/A” for all studies, as individual participant data allows for testing at all thresholds/cut-offs.
3. **Overall risk of bias:** Rate this item as “low” for all studies since the interpretation of the index test is fully automated in scoring self-report depressive symptom questionnaires and the individual participant data allows for testing at all thresholds/cut-offs.

4. **Applicability concerns:** Code “low” if the standard language version of the index test was used or if a translated version was used with an appropriate translation and back-translation process, or a translated version is located online. Code “unclear” if a translated version was used and it is not clear what steps were taken to ensure the quality of the translation or if only forward translation was used.

Domain 3: Reference Standard

1. **Signalling question 1 – Is the reference standard likely to correctly classify the condition?:** This question will be coded as “yes” for all studies because the use of a validated semi- or fully structured psychiatric interview to assess participants for a DSM or ICD diagnosis of MDD/MDE is an eligibility requirement.
2. **Signalling question 2 – Were the reference standard results interpreted without knowledge of the results of the index test?:** Code as “yes” if the person administering the diagnostic interview was blinded to the participant’s score on the index test, or if the diagnostic interview was administered before the index test. Code as “no” if the person administering the diagnostic interview was not blinded or was aware of the participant’s score on the index test. Code as “unclear” if the study does not indicate whether blinding occurred and we cannot ascertain whether blinding occurred.
3. **Signalling question 3 – Did a qualified person administer the reference standard?:** For structured clinical interviews, this will typically be coded “yes” as no specific clinical training is required. For semi-structured interviews, this will be coded “yes” if a trained mental health diagnostician administered the clinical interview (e.g., psychiatrist, psychologist, clinician, social worker, general practitioner, psychiatric nurse) or if non-clinicians who have comprehensive diagnostic experience and documented adequate training administered the clinical interview (e.g. trained doctoral student, research assistant, nurse, nurse practitioner, advanced practice nurse). Code “no” if individuals without the required training administered the reference standard (e.g., student, research assistant, nurse without documented extensive training necessary). Code “unclear” if the characteristics of personnel who administered the diagnostic interview cannot be ascertained or if a vague description of training is provided (e.g., trained research assistants with no additional information). If the name of the interviewer is provided in the article, but no credentials are listed, then code based on credentials retrieved online for the interviewer.
Fully structured: CIDI, DIS, MINI, CIS-R
Semi-structured: SCID, SCAN, DISH, CIS
4. **Overall risk of bias:** The coding of this item should consider blinding of the person administering the diagnostic interview to the participant’s score on the index test and the qualifications of individuals administering the reference standard interview.
5. **Applicability concerns:** This item will be coded as “low” for most standard language studies, since the use of a validated semi- or fully structured psychiatric interview to assess participants for a DSM or ICD diagnosis of MDD/MDE is an eligibility requirement. For translated versions of a validated reference standard, code “low” if a translated version was used with an appropriate translation and back-translation process. Code “unclear” if a translated version was used and it is not clear what steps were taken to ensure the quality of the translation or if only forward translation was used.

Domain 4: Flow and Timing

1. **Signalling question 1 – Was there an appropriate interval between index test and reference standard?:** Only patient data with two weeks or less between the index test and reference standard are included. Thus, code “yes” if index test and reference standard were administered within a week of each other. Code “unclear” if the period was greater than one week (but less than two weeks) or if the timing cannot be ascertained beyond knowing that it was < 2 weeks. Note that this item may be coded differently for different patients from the same study. Please see aggregated study information sheet to code this.

2. **Signalling question 2 – Did all patients receive a reference standard?:** This will typically be coded “yes”. If a portion of positive and negative screens receive the reference standard, and the patients selected were chosen randomly, code “yes”. If non-random selection based on clinical factors or the index test determined whether or not patients received a reference standard, then code “unclear” or “no”. An example of all patients not receiving a reference standard would occur, for instance, if patients who endorsed suicidality on the index test were referred for evaluation and did not receive the reference standard interview.
3. **Signalling question 3 – Did all patients receive the same reference standard?:** This question will typically be coded as “yes” for all studies, since the reference standard is almost always consistent within each study.
4. **Signalling question 4 – Were all patients included in the analysis?:** When coding for this question, compare the number of participants who received the index test to the number of participants who received the reference standard. Code as “yes” if at least 90% of participants who received the index test also received the reference standard, or vice versa, and were included in analyses. Code as “unclear” if this difference is $\geq 80\%$, but $< 90\%$ or if it cannot be determined. Code as “no” if it is $< 80\%$. If the study used randomly selected patients for either the index test or the reference standard, do not count the participants who did not receive the reference standard for that reason as missing. In “Notes”, please provide the relevant numbers and percentages used to make a determination.
5. **Overall risk of bias:** Rate as “low”, “high”, or “unclear” risk of bias. Given that questions 2 and 3 will typically be coded as “yes”, use the following rules to code the overall risk of bias:

SQ1 = UNCLEAR and SQ4 = YES: code as UNCLEAR risk of bias

SQ1 = UNCLEAR and SQ4 = UNCLEAR: code as UNCLEAR risk of bias

SQ1 = UNCLEAR and SQ4 = NO: code as HIGH risk of bias if the % in SQ4 is $< 50\%$ and code as UNCLEAR risk of bias if the % in SQ4 is $\geq 50\%$

SQ1 = YES and SQ4 = UNCLEAR: code as UNCLEAR risk of bias

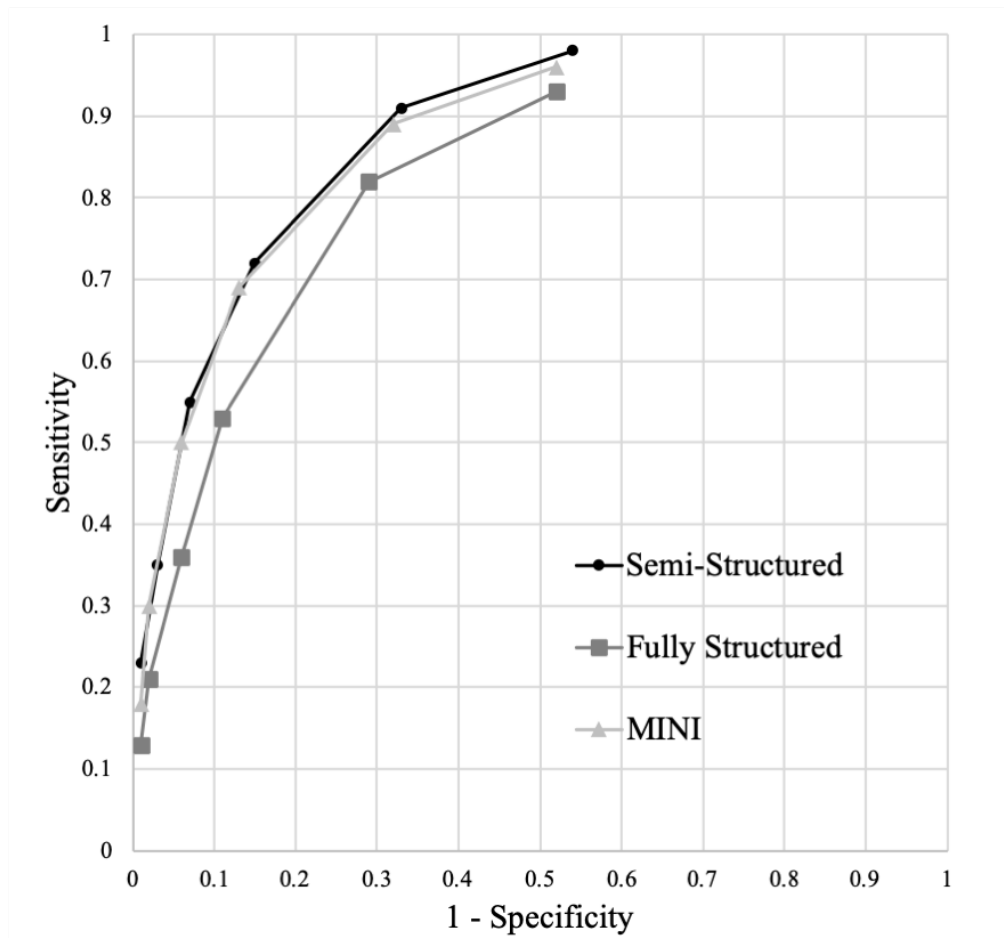
SQ1 = YES and SQ4 = YES: code as LOW risk of bias

SQ1 = YES and SQ4 = NO: code as HIGH risk of bias if the % in SQ4 is $< 50\%$ and code as UNCLEAR risk of bias if the % in SQ4 is $\geq 50\%$

Note: If “IPD” was selected for signalling question 1, and the overall risk of bias rating depends on the individual patient rating in signalling question 1, then rate as “IPD” and indicate which participants should receive which bias rating (for example, participants administered the reference standard within 1 week are rated as “low”, whereas those administered the reference standard within 1-2 weeks are rated as “unclear”).

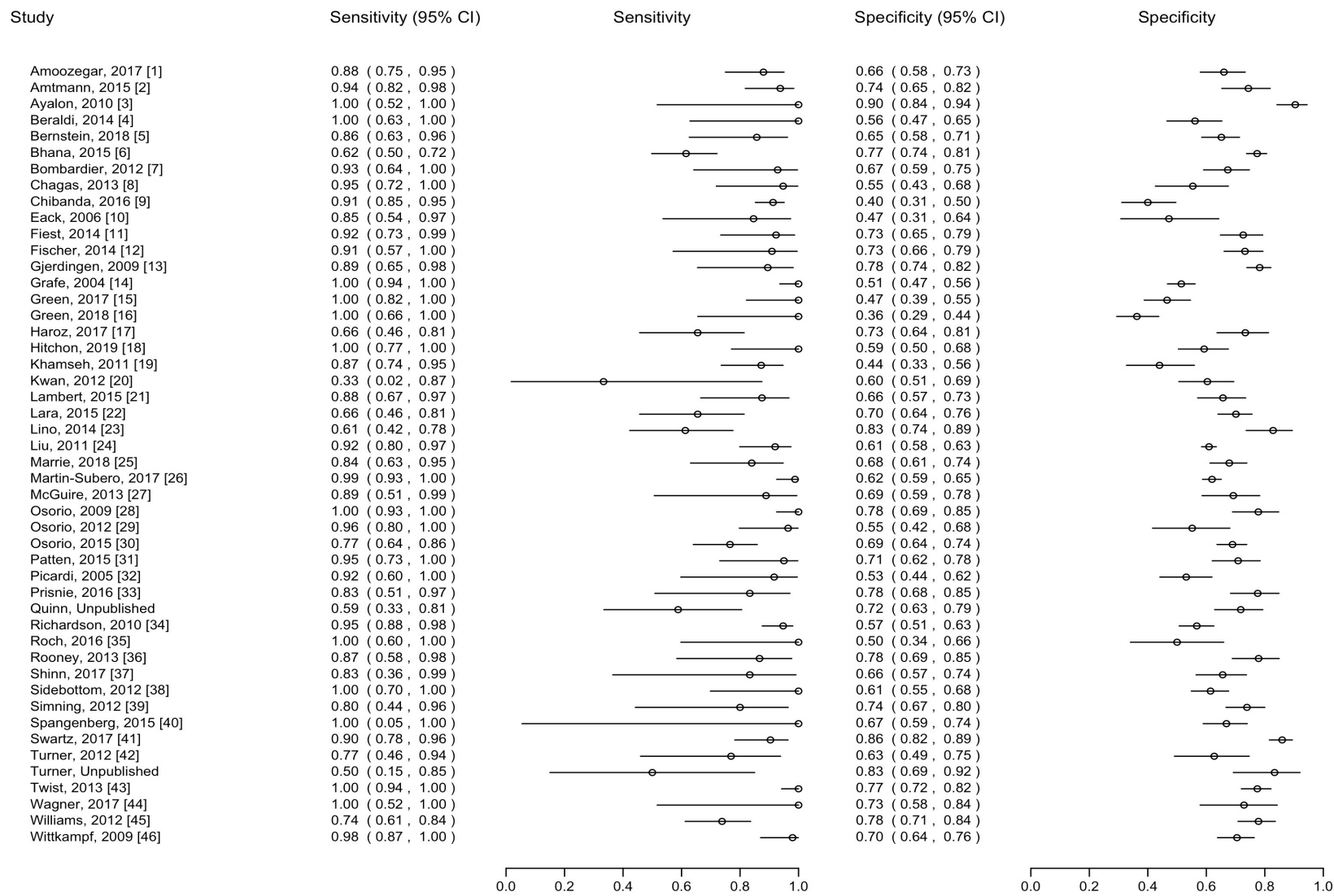
Please indicate factors in decision in “Notes”.

eFigure 1. Receiver operating characteristic (ROC) plots comparing sensitivity and specificity estimates for Patient Health Questionnaire-2 (PHQ-2) cutoffs 1-6 among semi-structured diagnostic interviews, fully structured diagnostic interviews, and the Mini International Neuropsychiatric Interview (MINI)^a

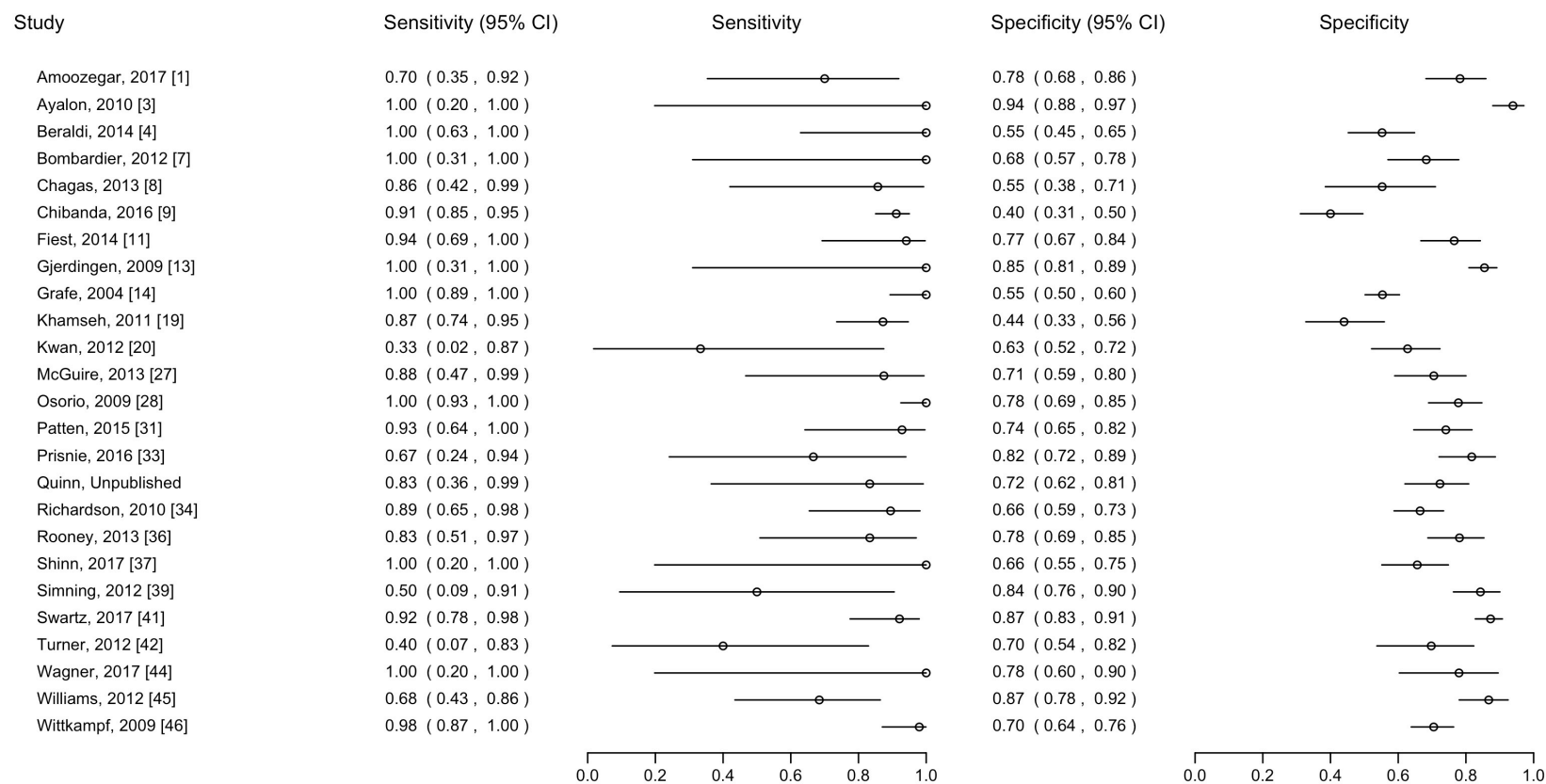


^aArea under the curve and 95% confidence intervals: 0.875 (0.864, 0.887) for semi-structured interviews, 0.821 (0.807, 0.835) for fully structured interviews, and 0.866 (0.854, 0.877) for MINI

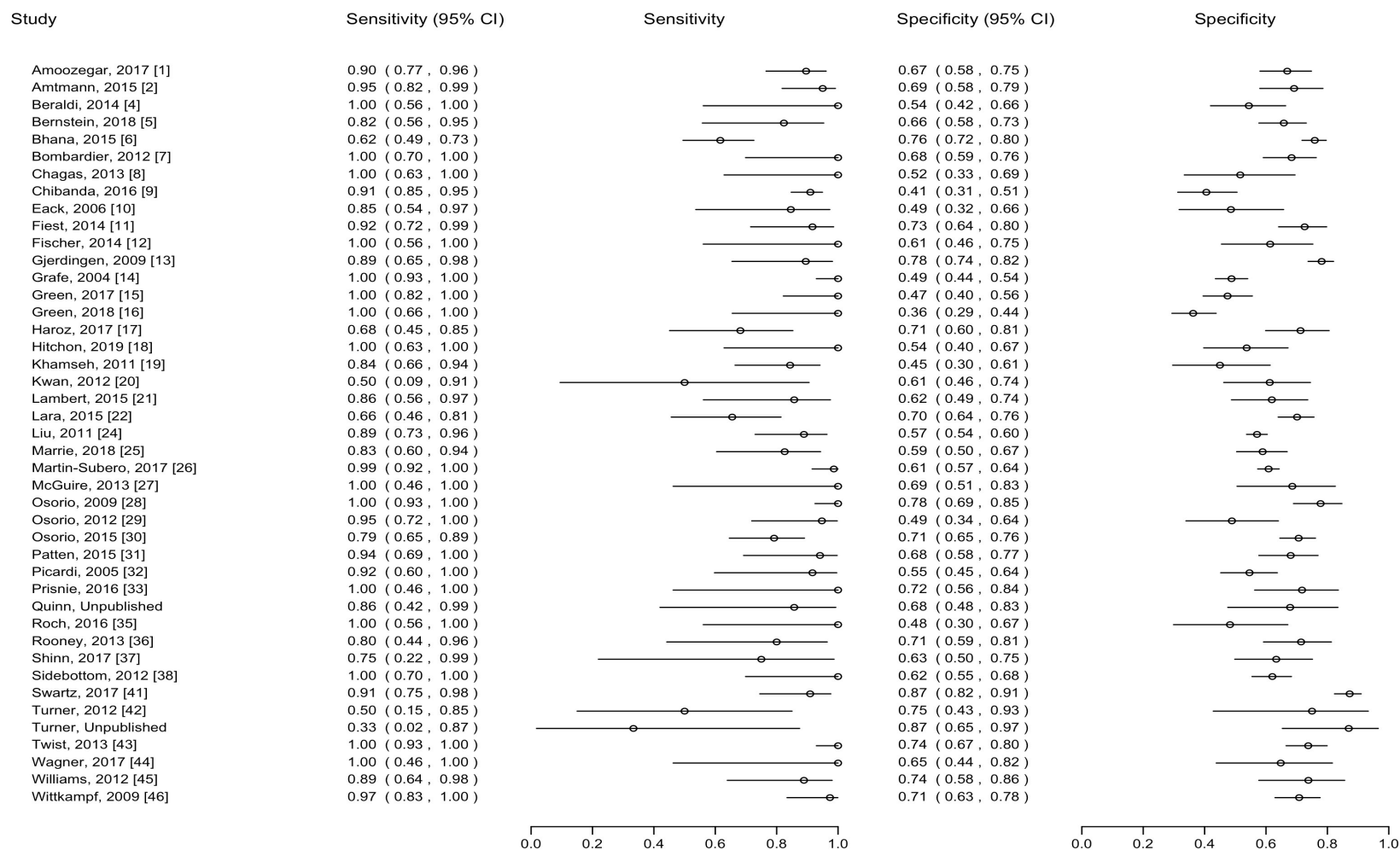
eFigure 2a. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 48; N Participants = 11,703; N major depression = 1,538)



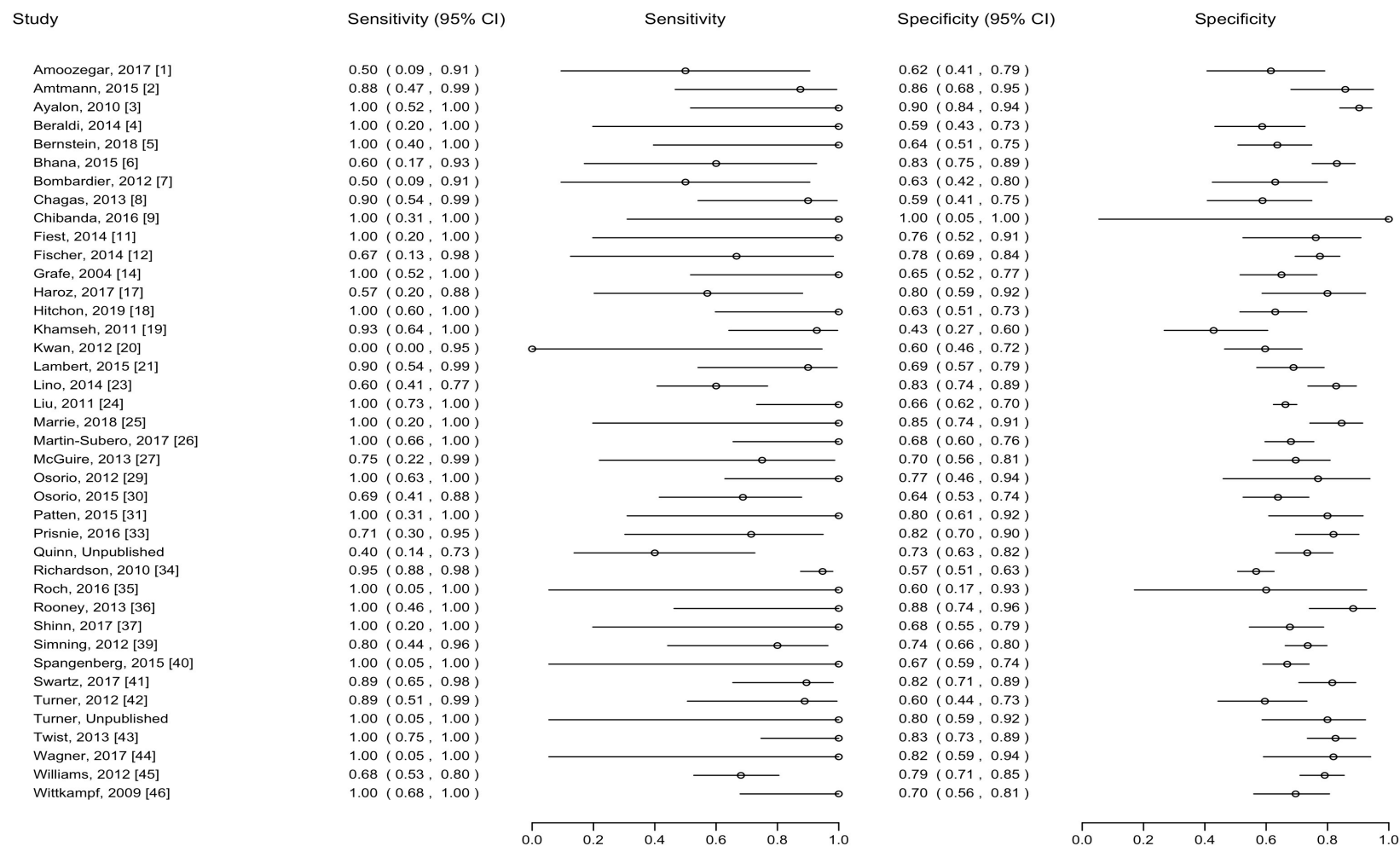
eFigure 2b. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants verified to not currently be diagnosed or receiving treatment for a mental health problem, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 25; N Participants = 3,708; N major depression = 527)



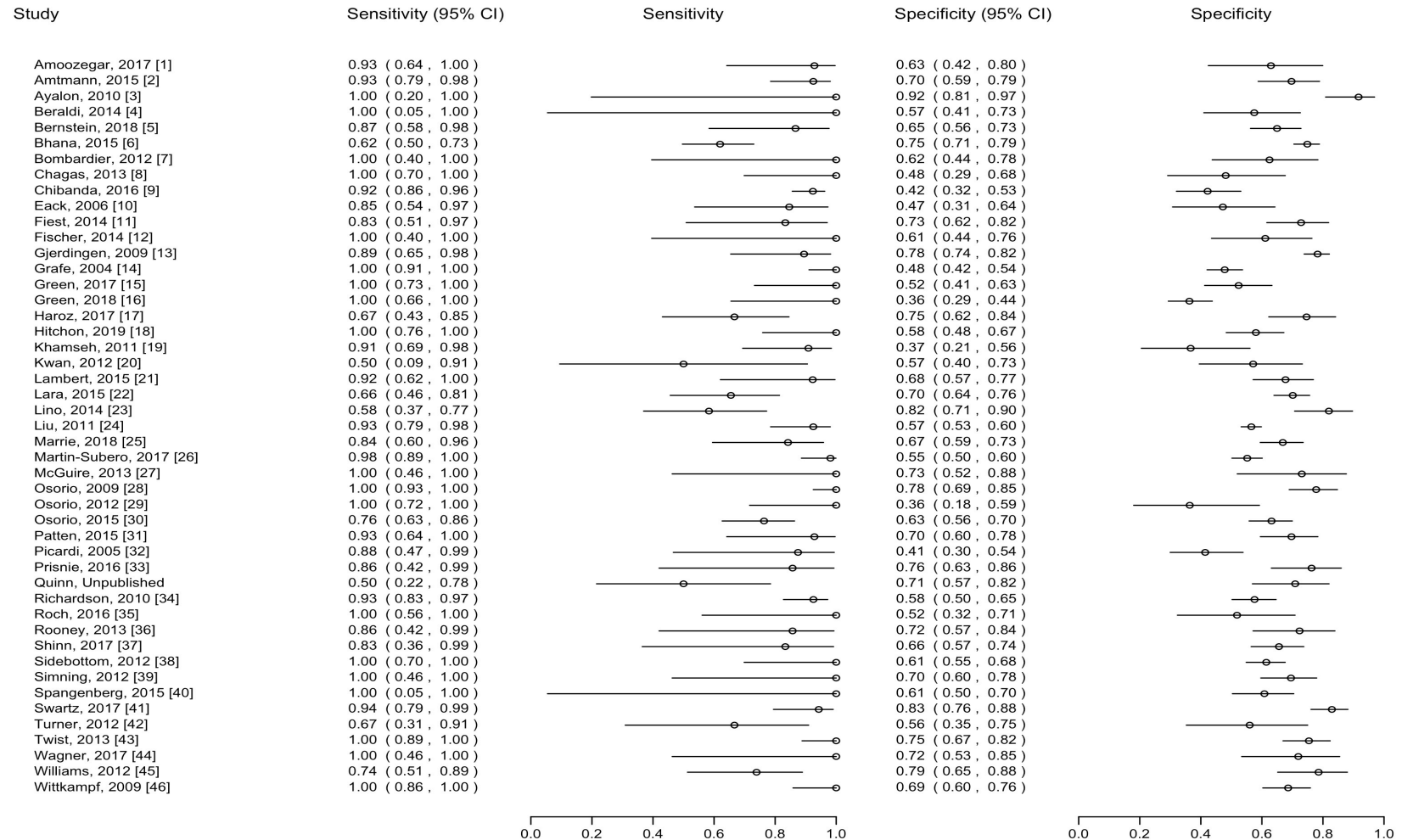
eFigure 2c. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants aged < 60, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 43; N Participants = 7,759; N major depression = 1,117)



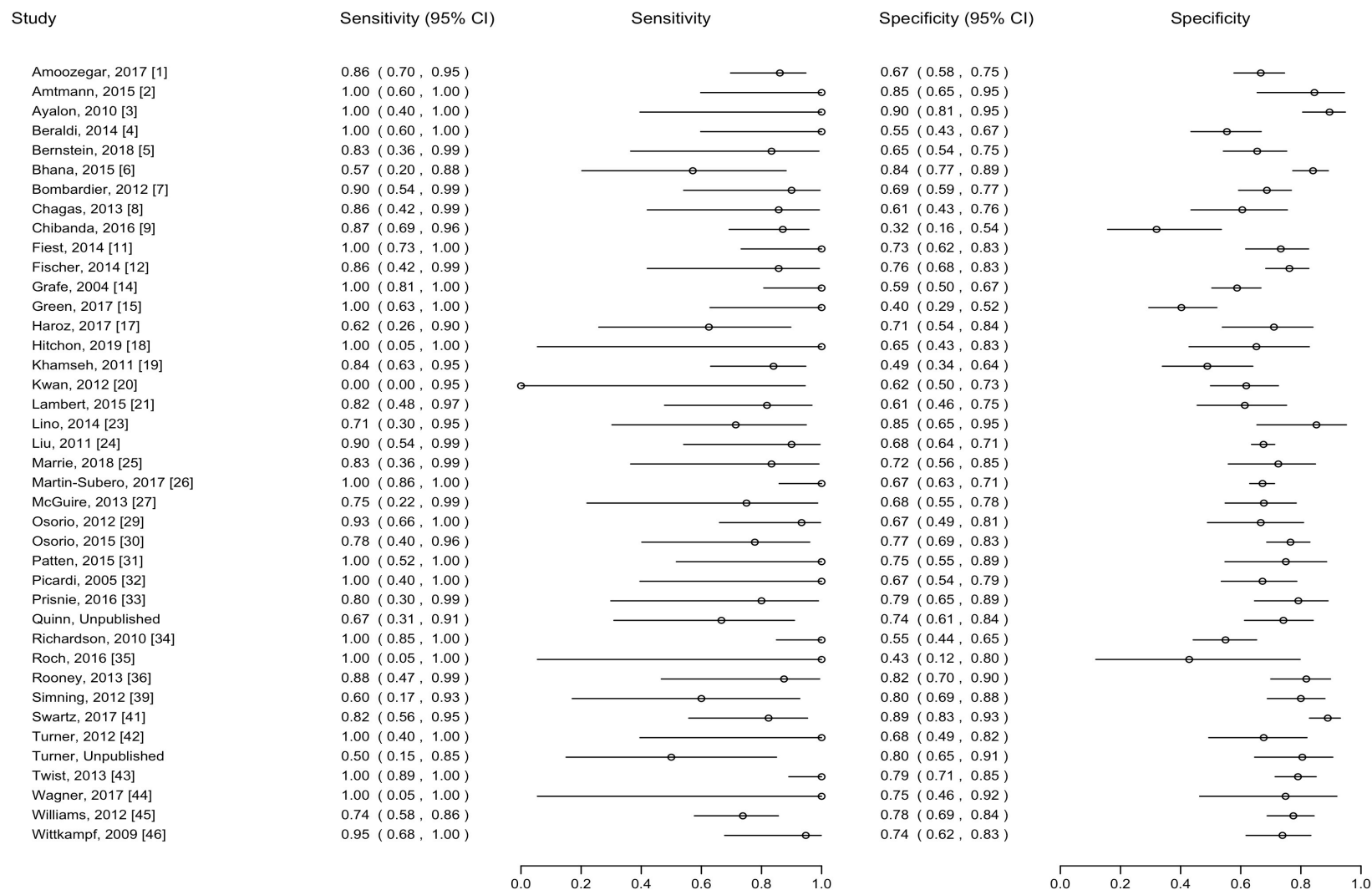
eFigure 2d. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants aged ≥ 60 , among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 40; N Participants = 3,875; N major depression = 415)



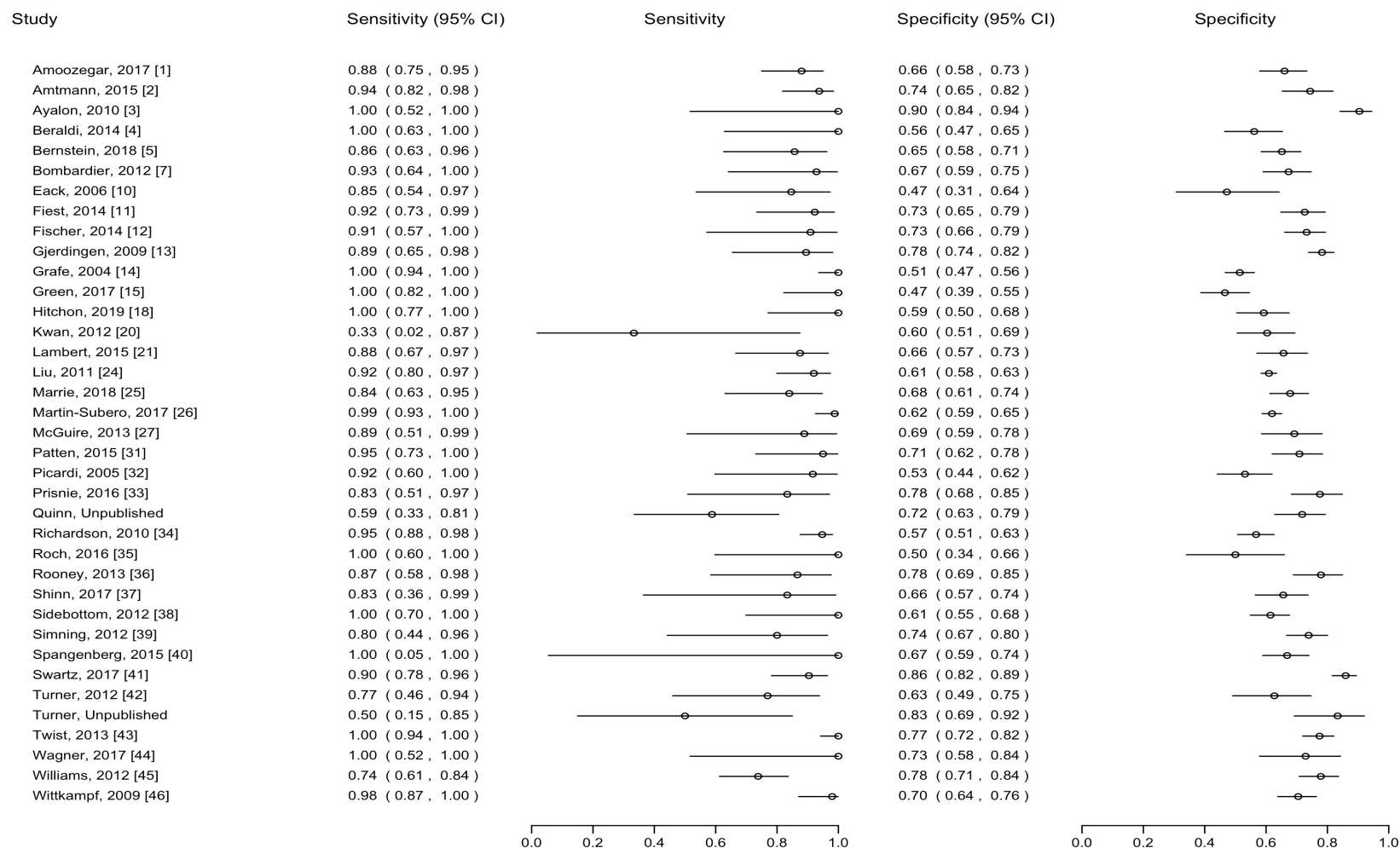
eFigure 2e. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among women, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 47; N Participants = 7,280; N major depression = 1,054)



eFigure 2f. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among men, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 40; N Participants = 4,345; N major depression = 484)

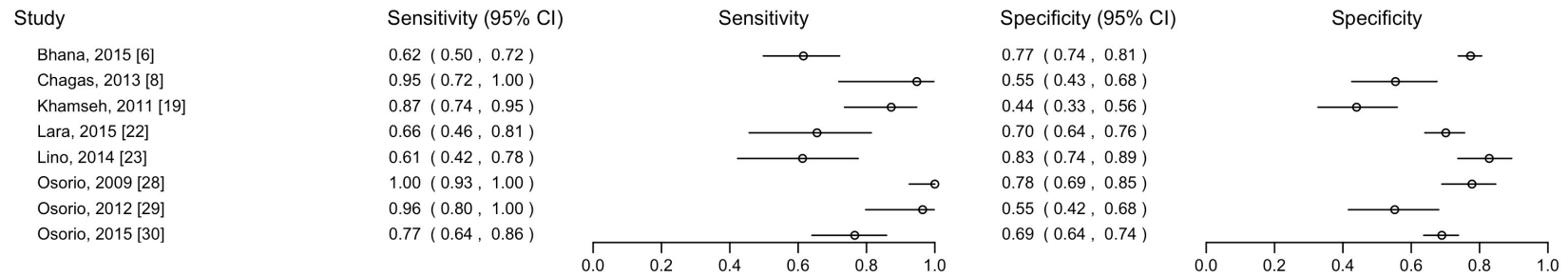


eFigure 2g. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a country with a very high human development index, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 37; N Participants = 9,156; N major depression = 994)

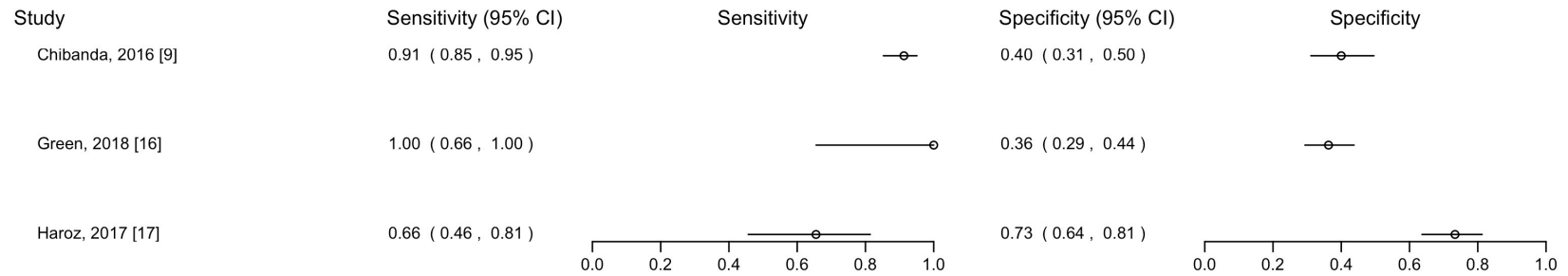


eFigure 2h. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a country with a high human development index, among studies that used a semi-structured diagnostic interview as the reference standard (N

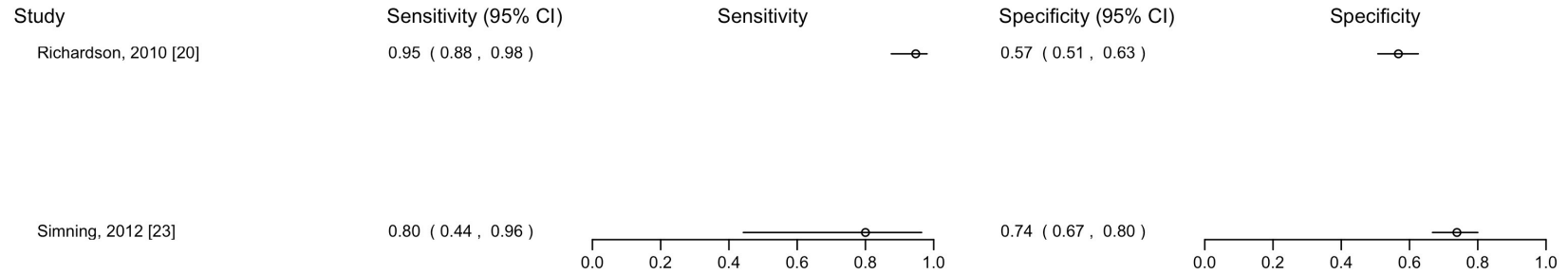
Studies = 8; N Participants = 1,957; N major depression = 356)



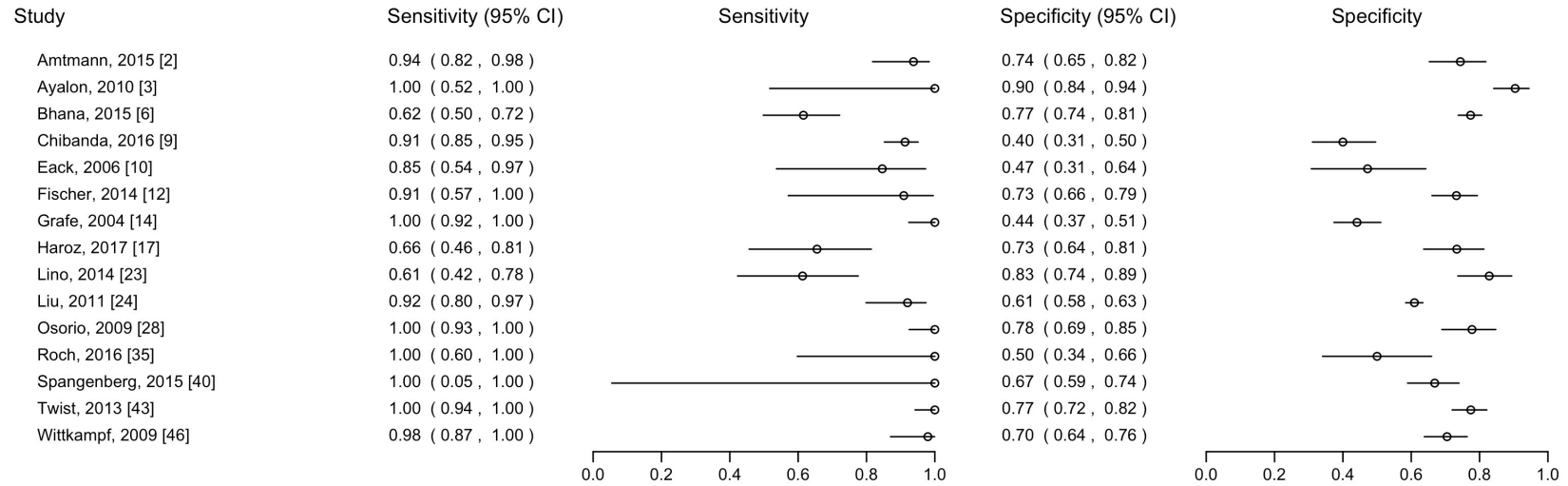
eFigure 2i. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a country with a low-medium human development index, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 3; N Participants = 590; N major depression = 188)



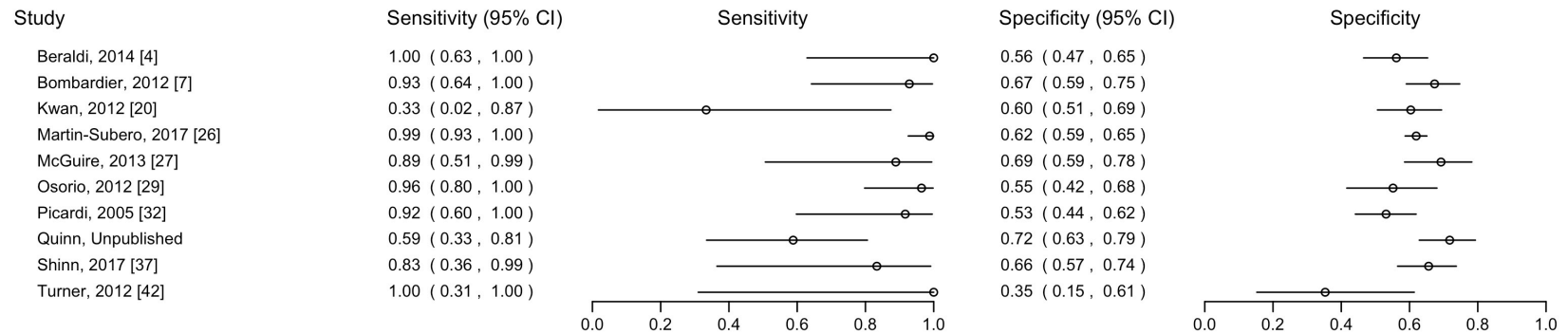
eFigure 2j. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a non-medical setting, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 2; N Participants = 567; N major depression = 105)



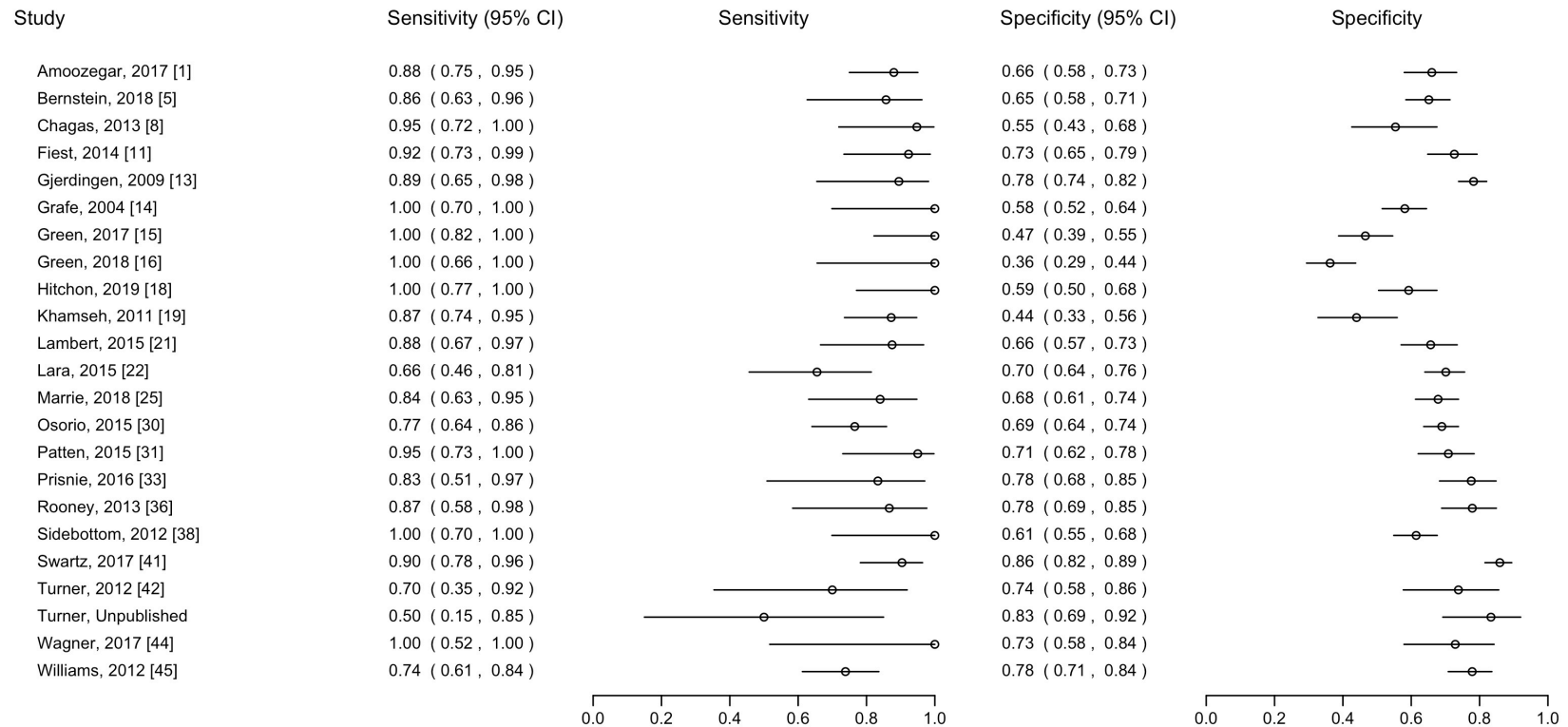
eFigure 2k. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a primary care setting, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 15; N Participants = 4,569; N major depression = 667)



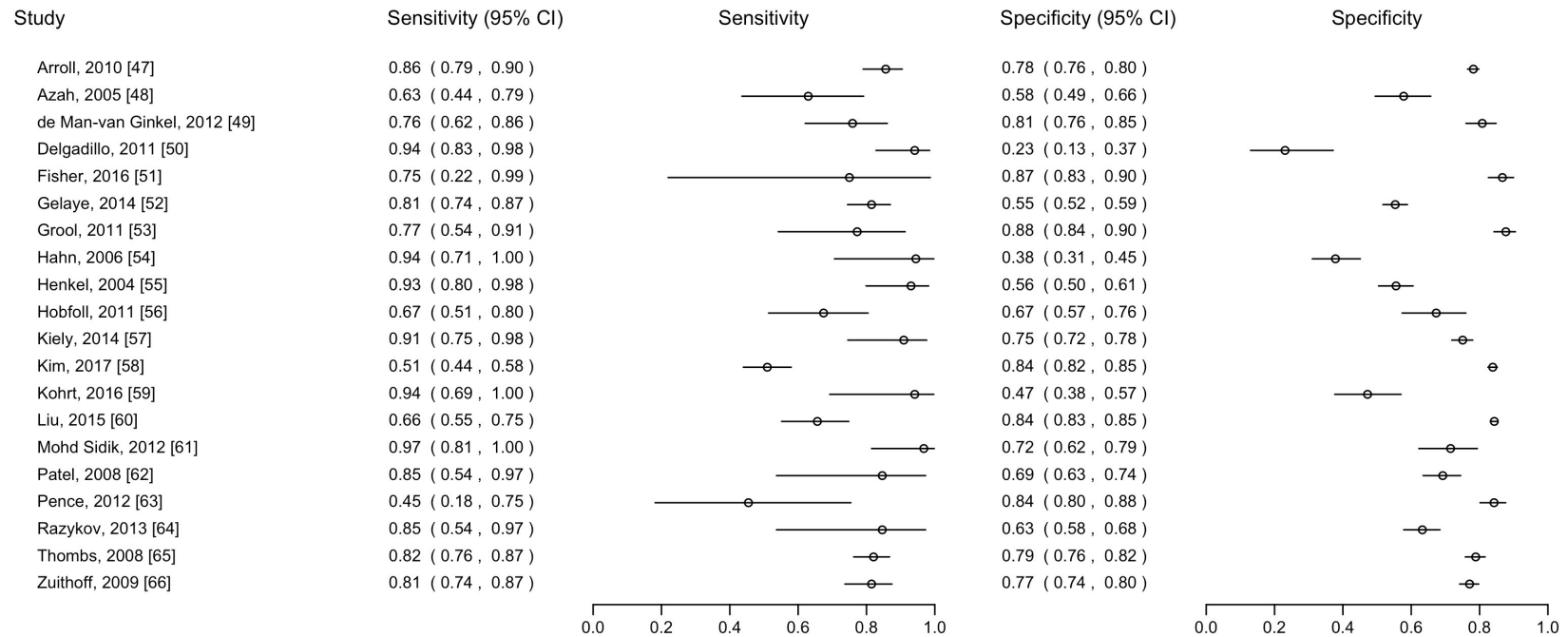
eFigure 2I. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from an inpatient specialty care setting, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 10; N Participants = 2,019; N major depression = 184)



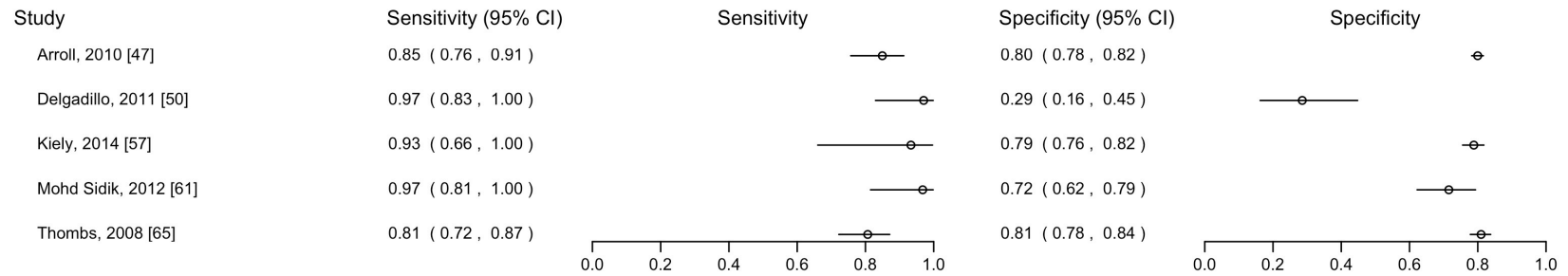
eFigure 2m. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from an outpatient specialty care setting, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 23; N Participants = 4,548; N major depression = 582)



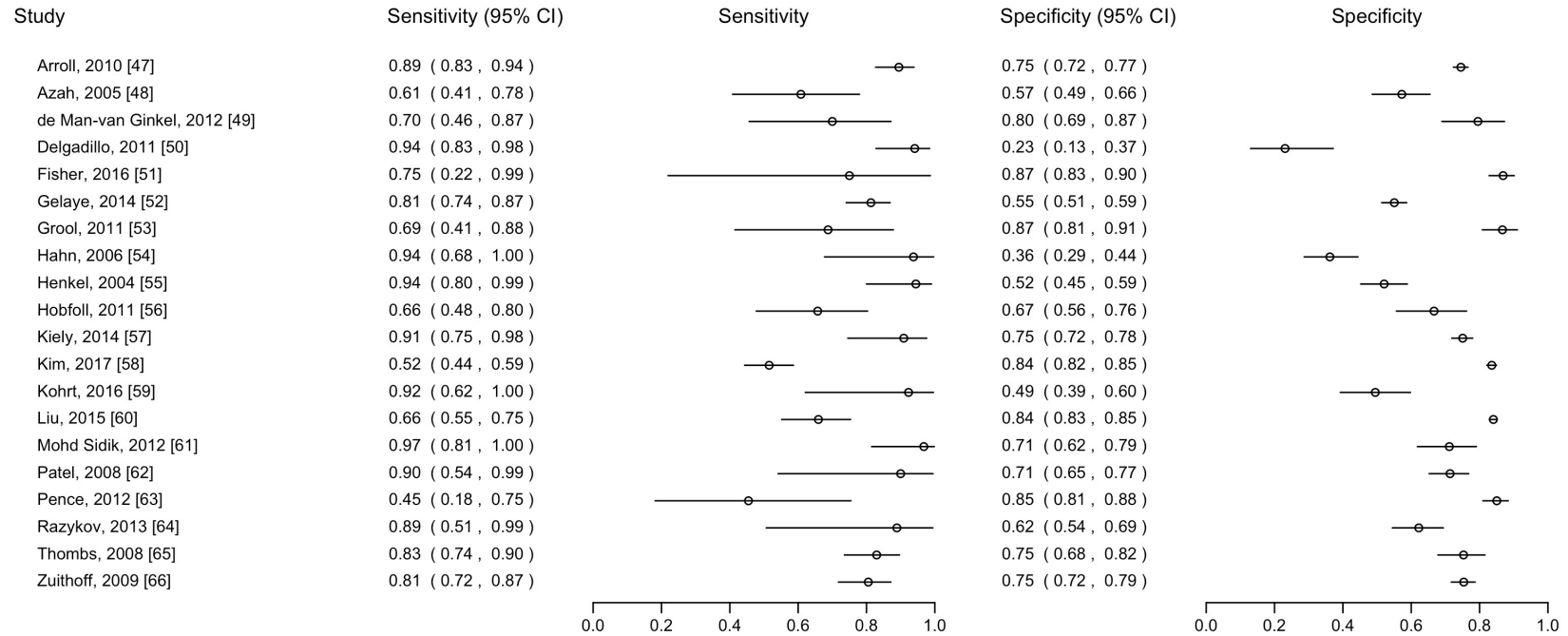
eFigure 2n. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 20; N Participants = 17,319; N major depression = 1,365)



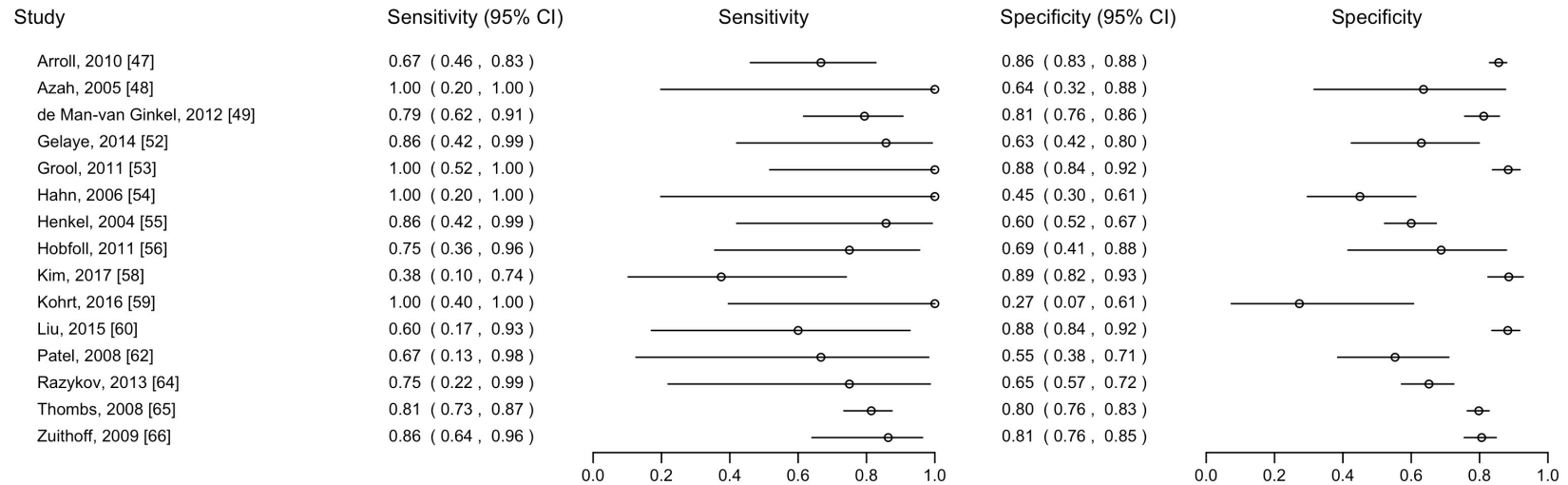
eFigure 2o. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants verified to not currently be diagnosed or receiving treatment for a mental health problem, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 5; N Participants = 4,050; N major depression = 292)



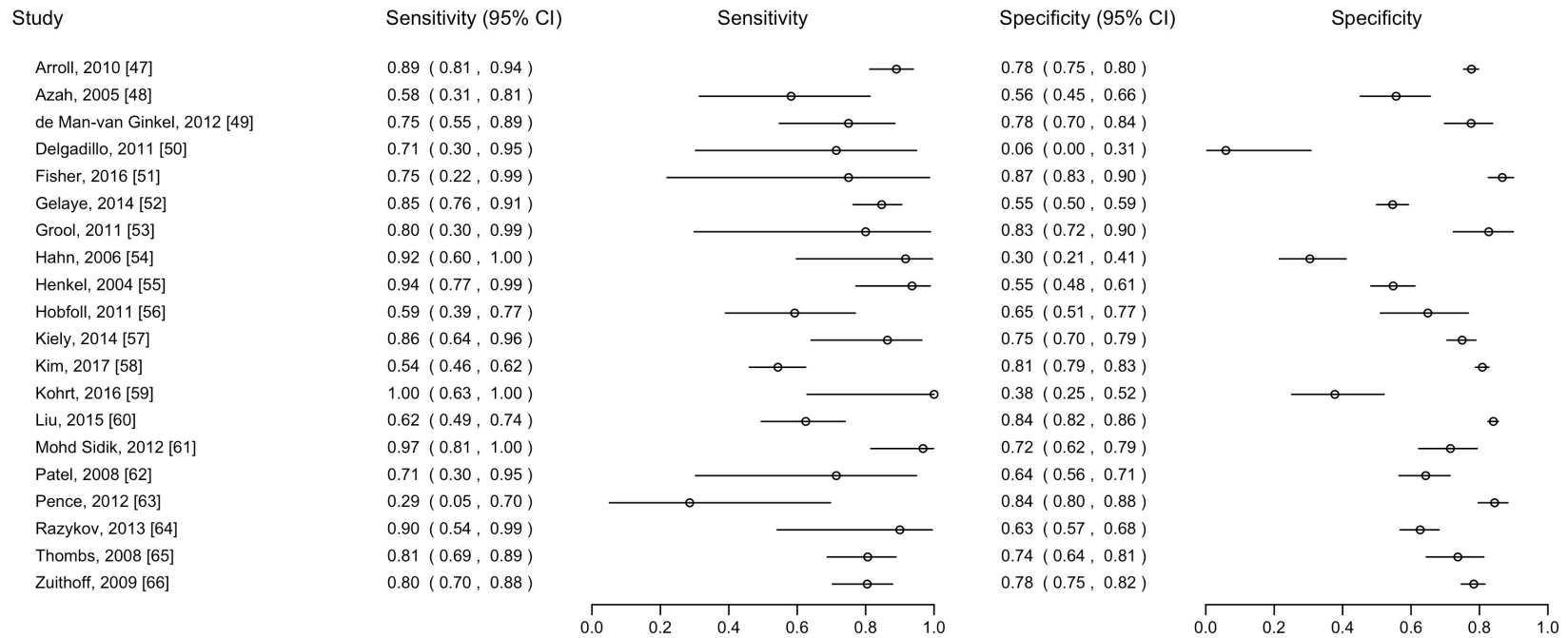
eFigure 2p. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants aged < 60, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 20; N Participants = 13,901; N major depression = 1,097)



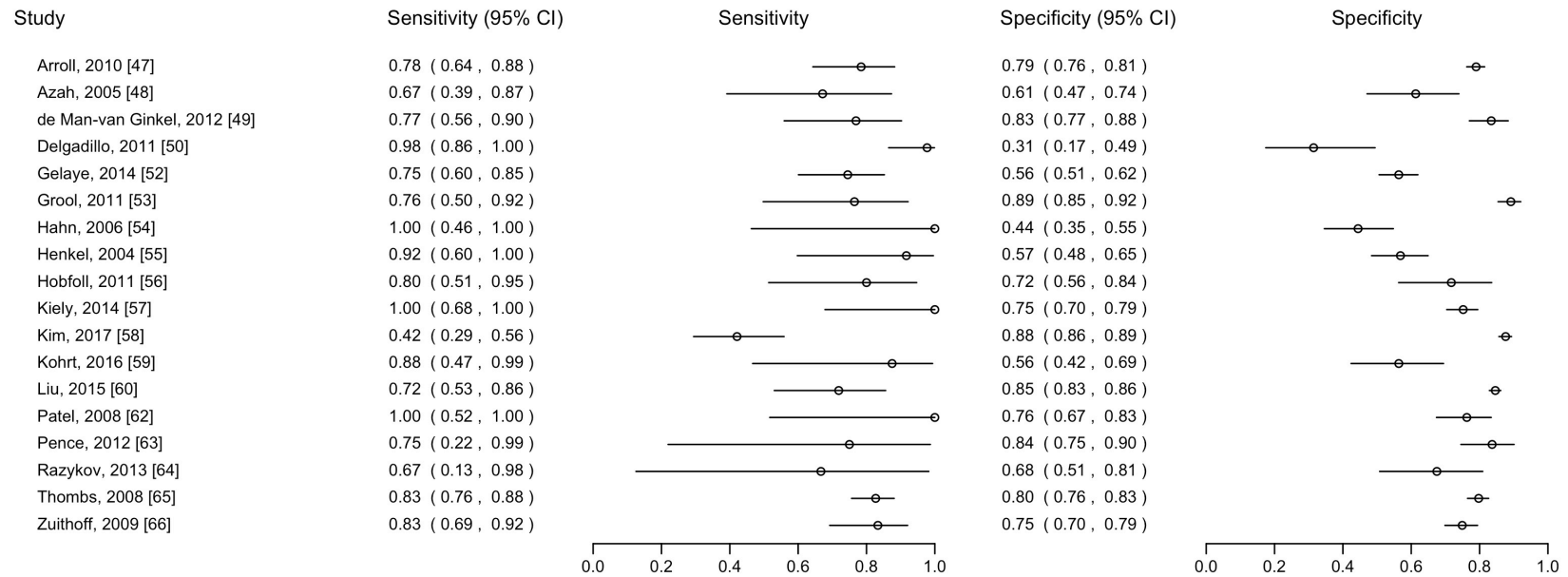
eFigure 2q. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants aged ≥ 60 , among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 15; N Participants = 3,400; N major depression = 268)



eFigure 2r. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among women, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 20; N Participants = 9,690; N major depression = 802)

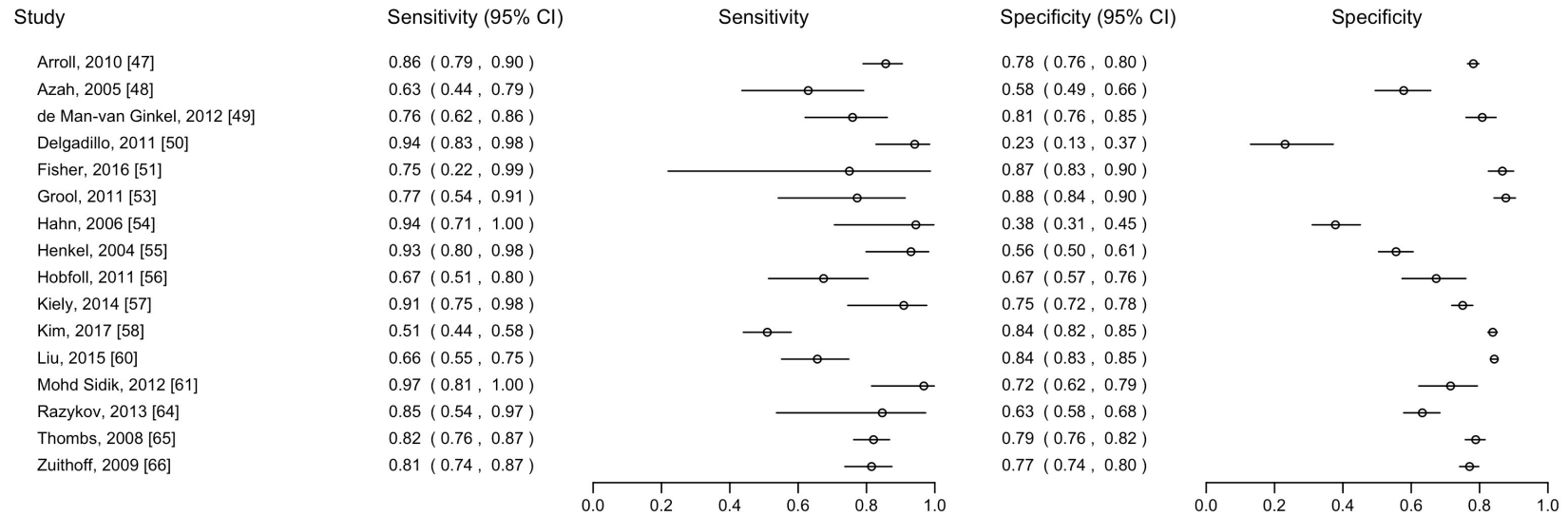


eFigure 2s. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among men, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 18; N Participants = 7,619; N major depression = 561)

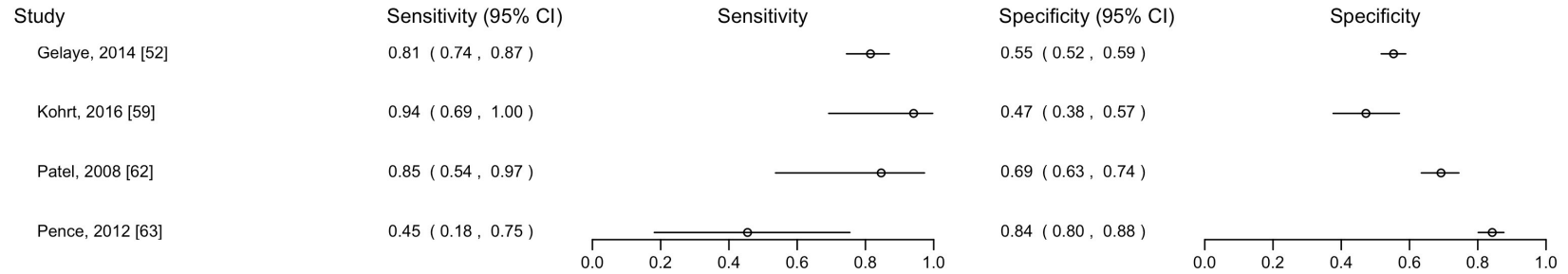


eFigure 2t. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a country with a very high human development index, among studies that used a fully structured diagnostic interview as the reference standard

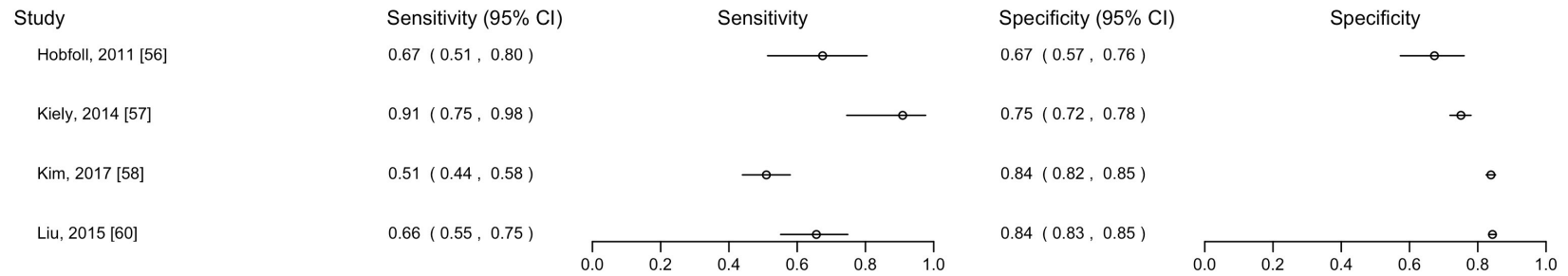
(N Studies = 16; N Participants = 15,574; N major depression =1,162)



eFigure 2u. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a country with a low-medium human development index, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 4; N Participants = 1,745; N major depression = 203)

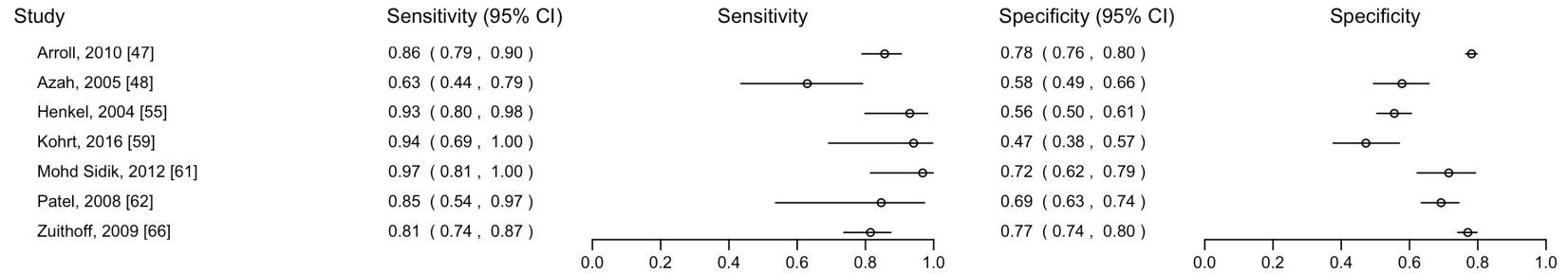


eFigure 2v. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a non-medical setting, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 4; N Participants = 8,316; N major depression = 378)

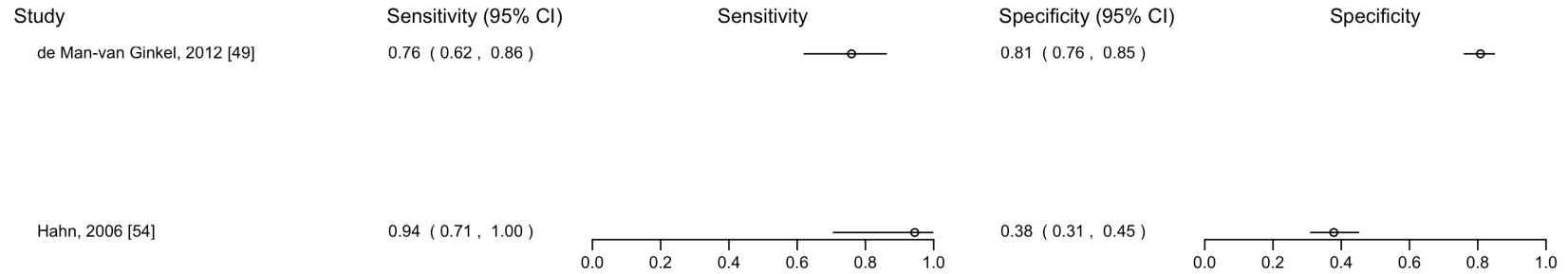


eFigure 2w. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a primary care setting, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 7; N

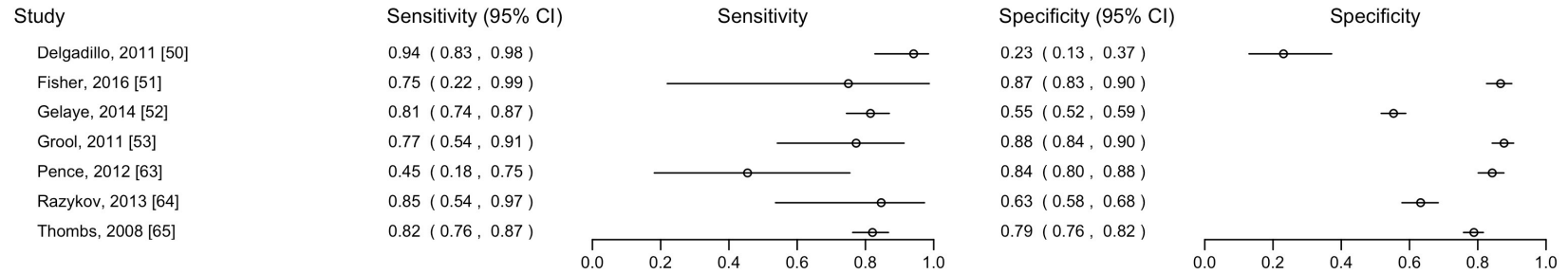
Participants = 4,789; N major depression = 429)



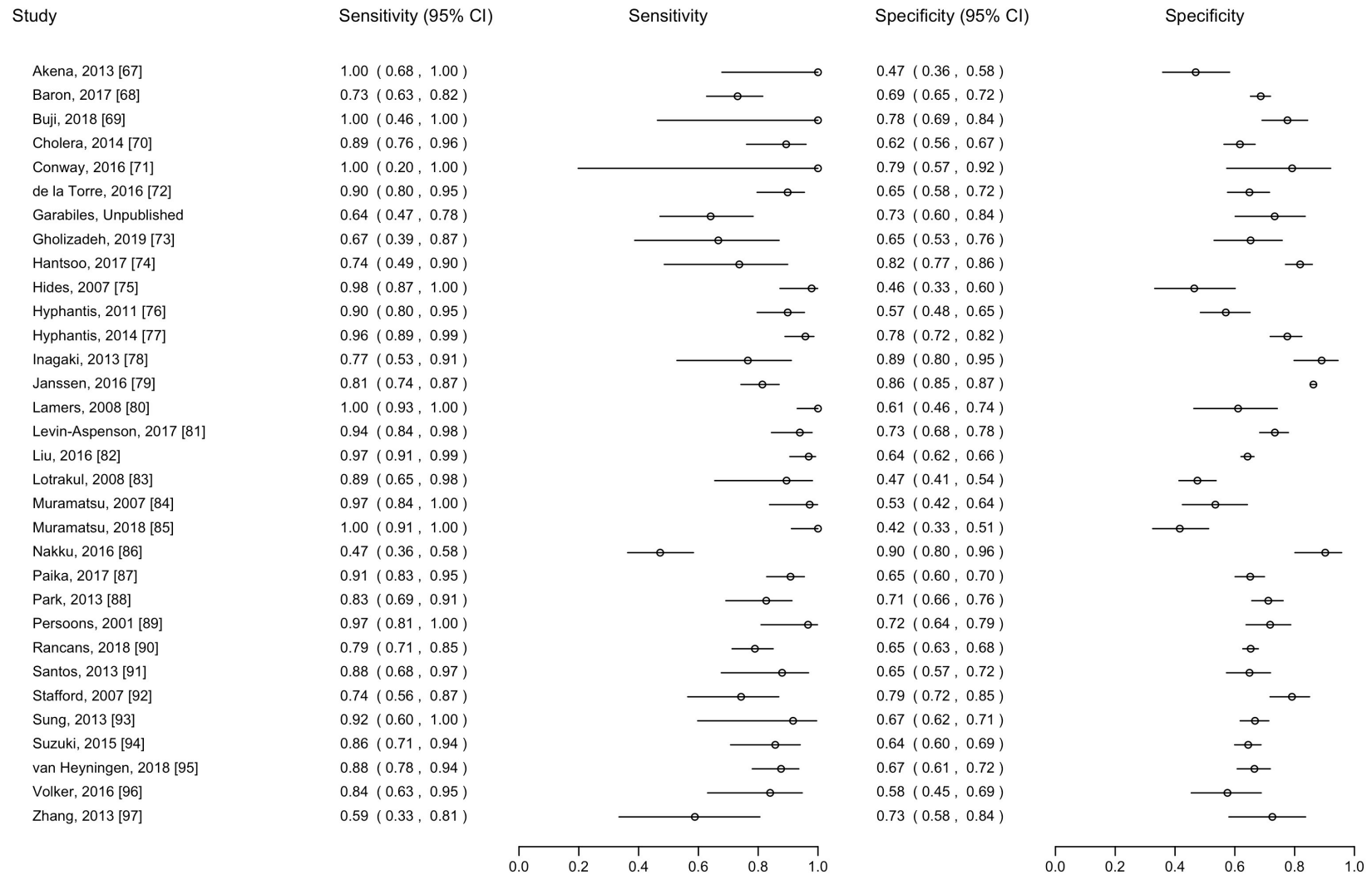
eFigure 2x. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from an inpatient specialty care setting, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 2; N Participants = 593; N major depression = 72)



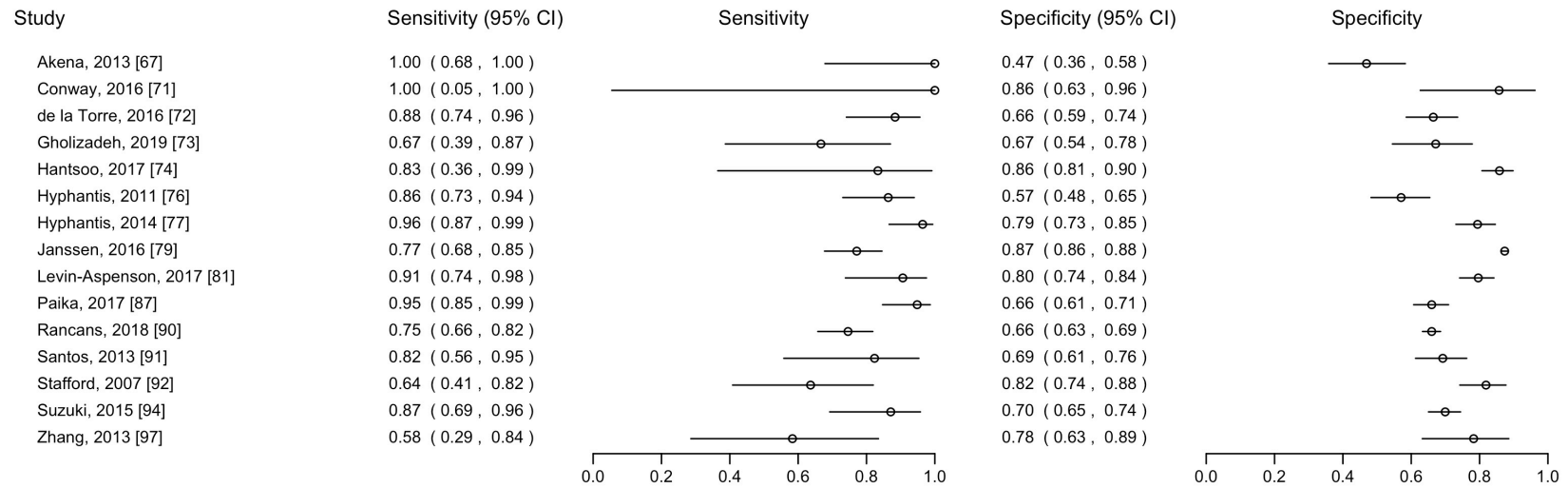
eFigure 2y. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from an outpatient specialty care setting, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 7; N Participants = 3,621; N major depression = 486)



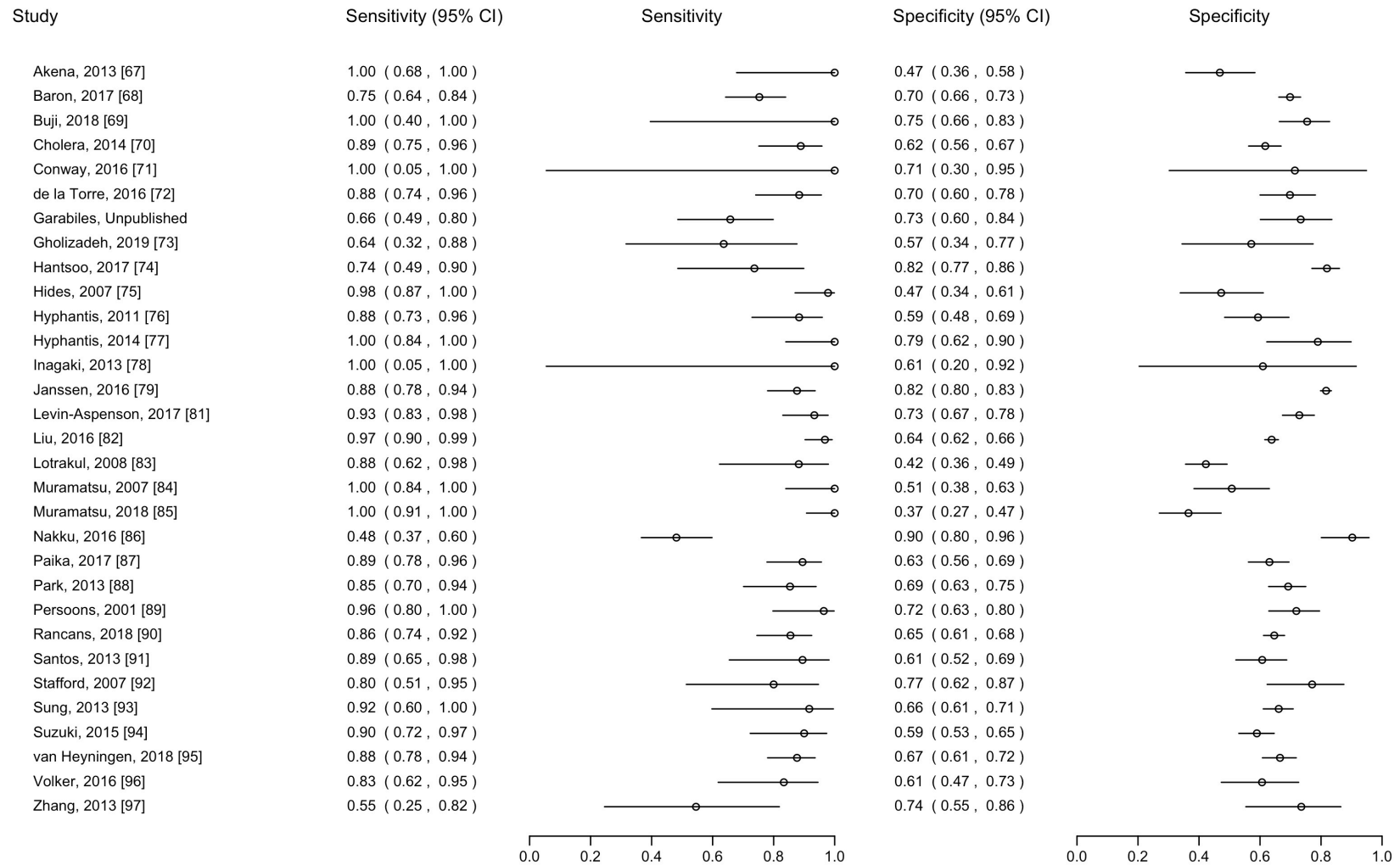
eFigure 2z. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2, among studies that used the MINI as the reference standard (N Studies = 32; N Participants = 15,296; N major depression = 1,669)



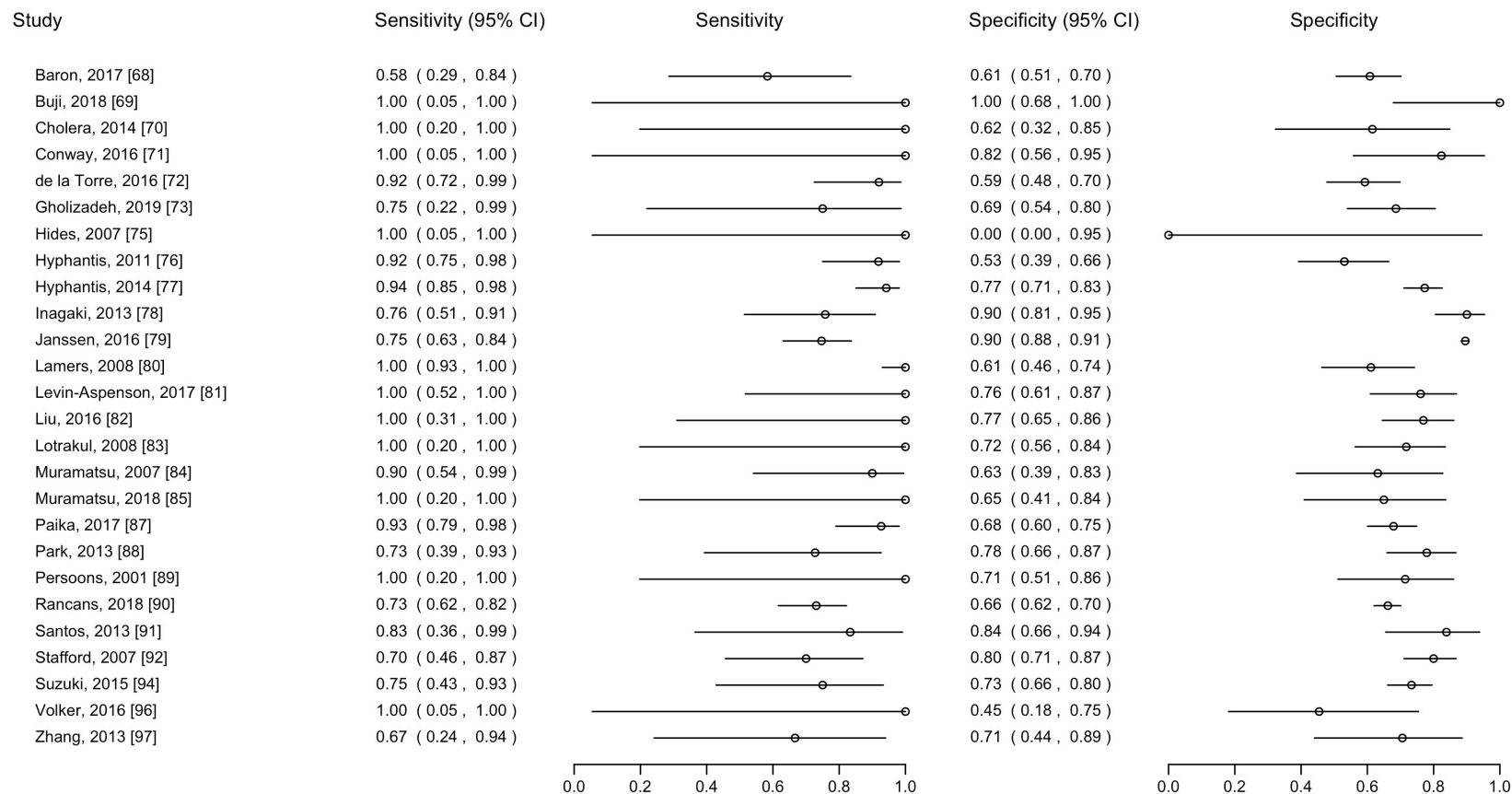
eFigure 2aa. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants verified to not currently be diagnosed or receiving treatment for a mental health problem, among studies that used the MINI as the reference standard (N Studies = 15; N Participants = 8,390; N major depression = 581)



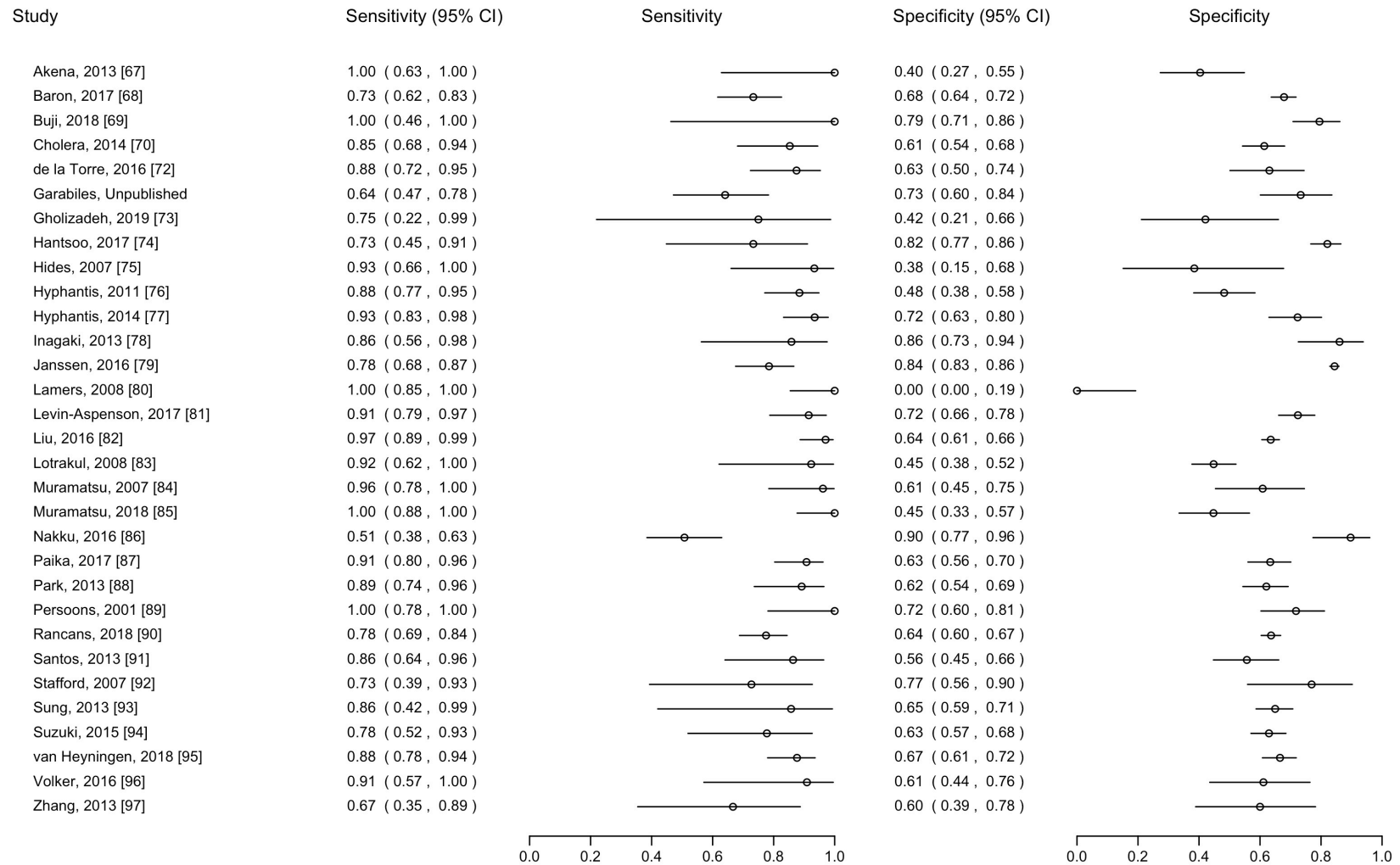
eFigure 2ab. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants aged < 60, among studies that used the MINI as the reference standard (N Studies = 31; N Participants = 10,071; N major depression = 1,153)



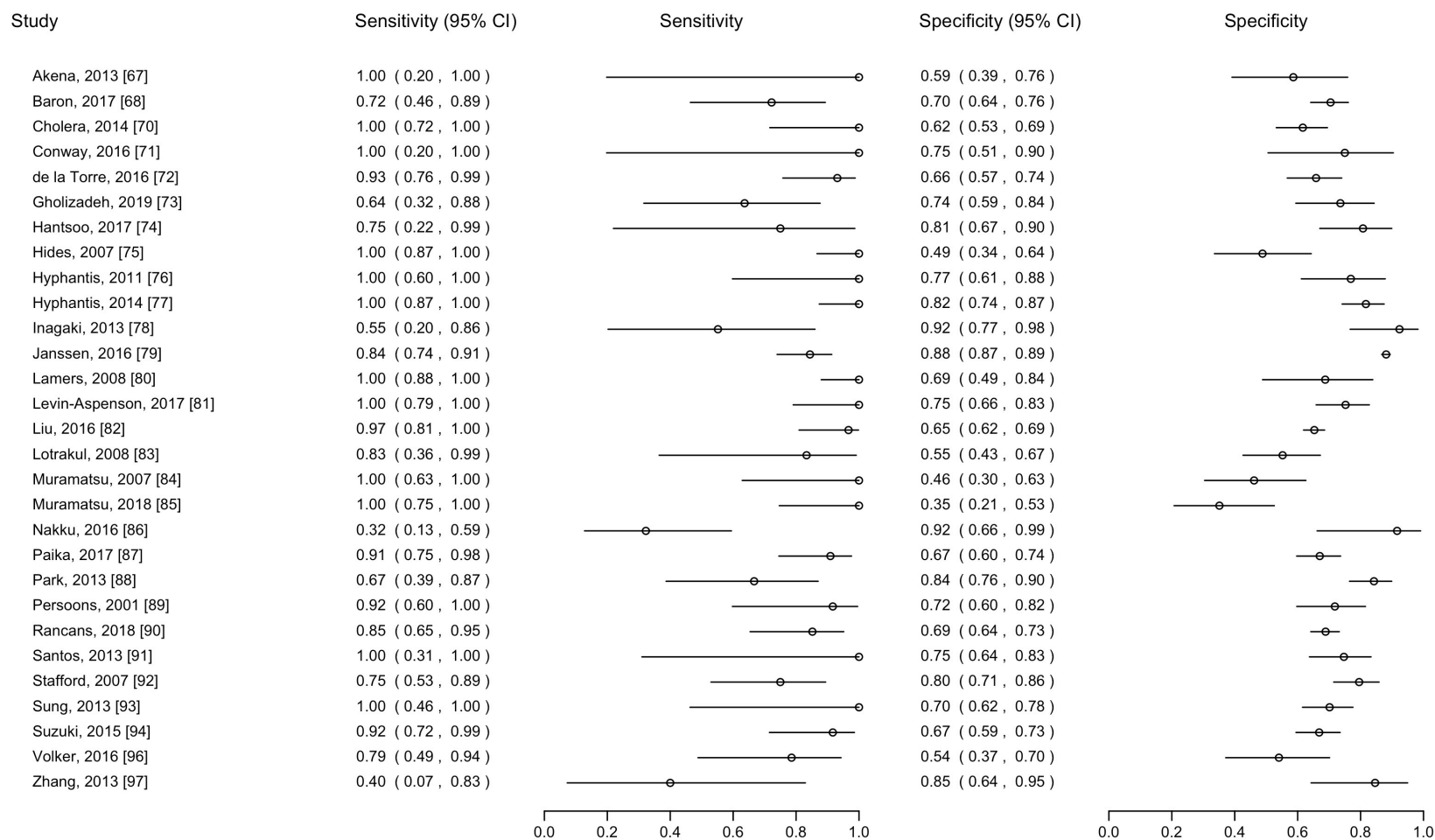
eFigure 2ac. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants aged ≥ 60 , among studies that used the MINI as the reference standard (N Studies = 26; N Participants = 5,192; N major depression = 506)



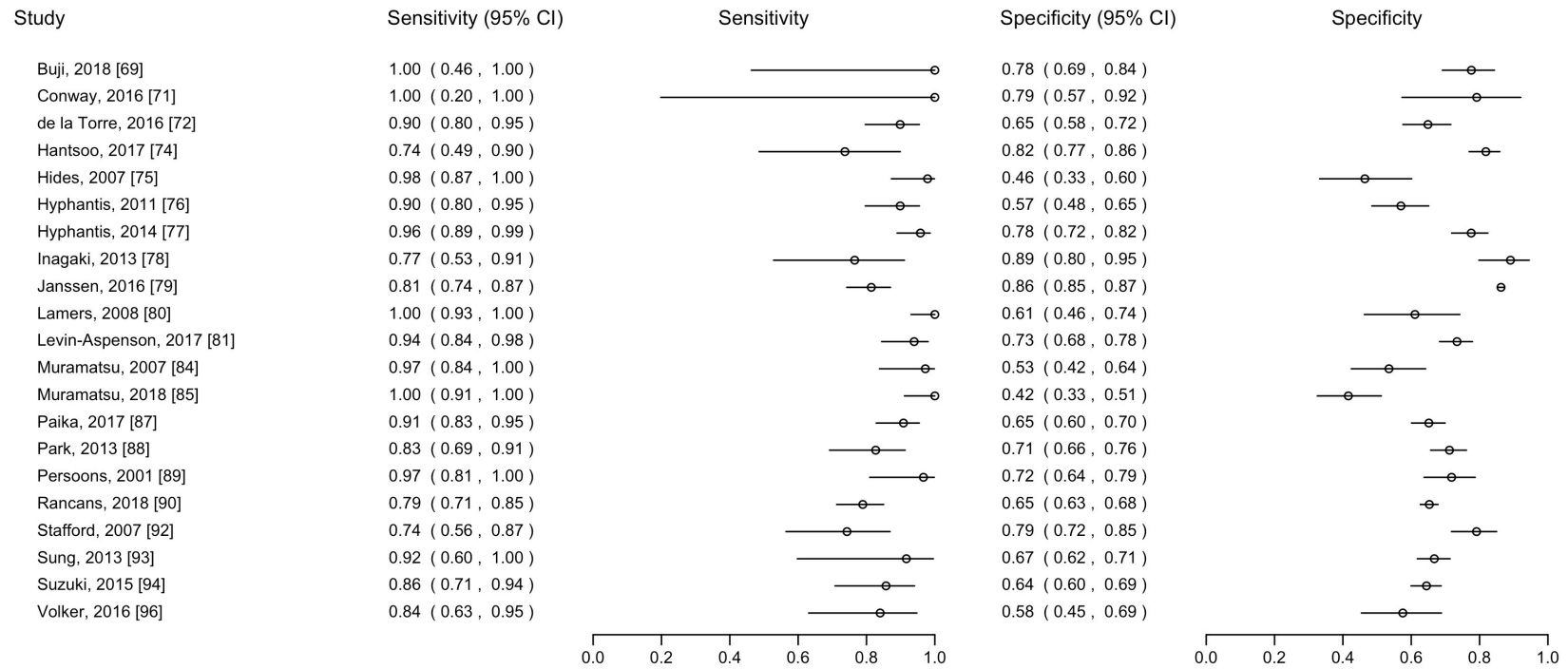
eFigure 2ad. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among women, among studies that used the MINI as the reference standard (N Studies = 31; N Participants = 9,053; N major depression = 1,138)



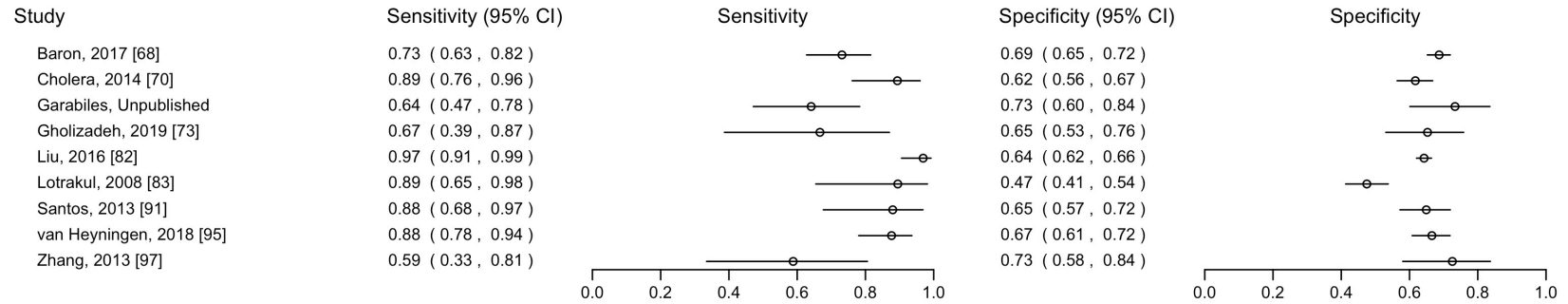
eFigure 2ae. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among men, among studies that used the MINI as the reference standard (N Studies = 29; N Participants = 6,225; N major depression = 530)



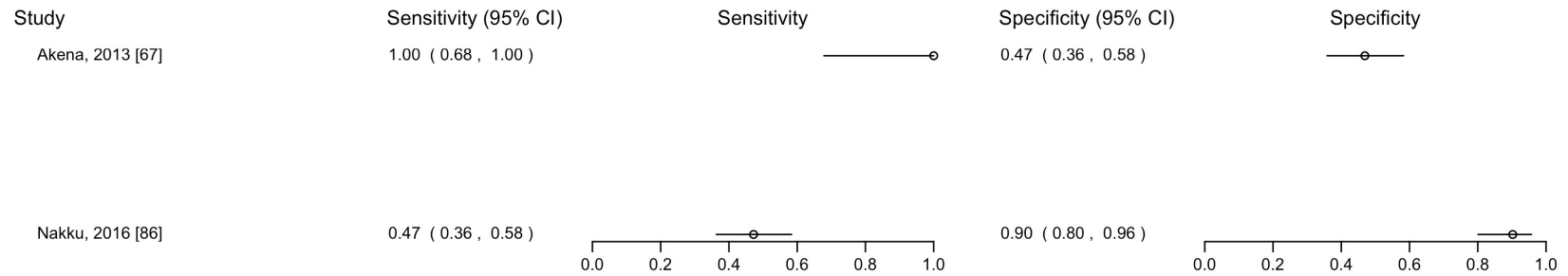
eFigure 2af. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a country with a very high human development index, among studies that used the MINI as the reference standard (N Studies = 21; N Participants = 10,699; N major depression = 1,141)



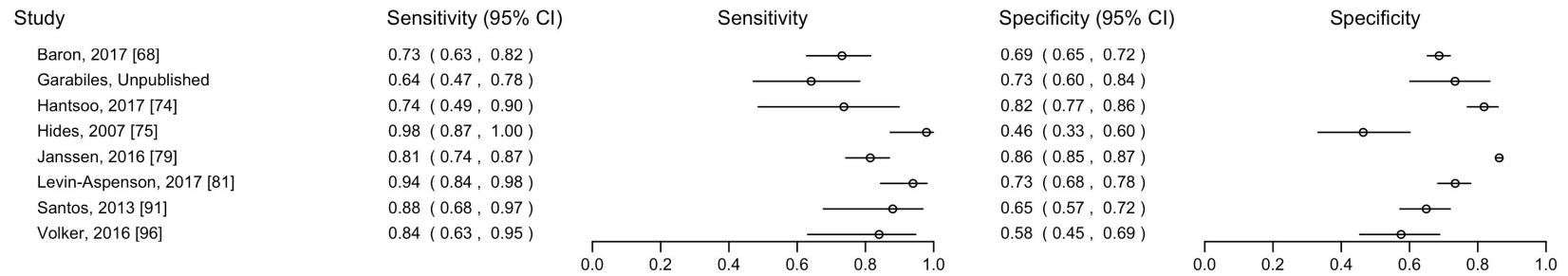
eFigure 2ag. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a country with a high human development index, among studies that used the MINI as the reference standard (N Studies = 9; N Participants = 4,352; N major depression = 433)



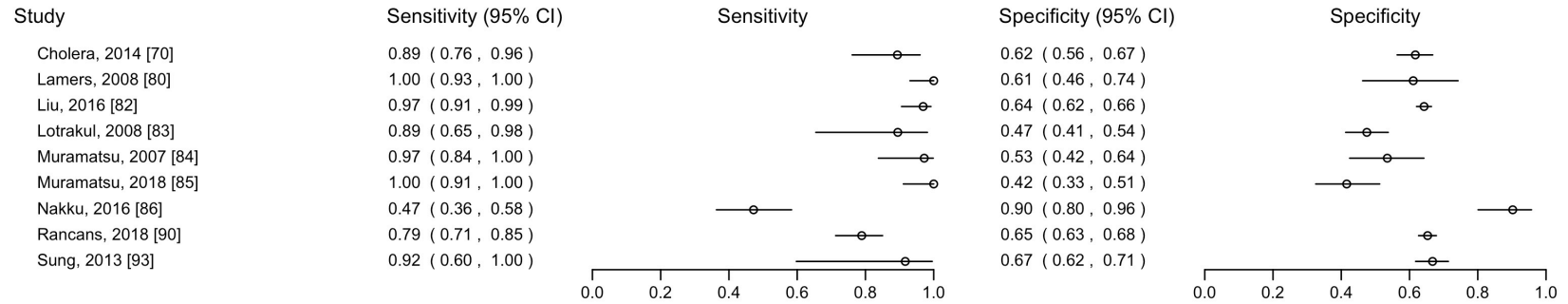
eFigure 2ah. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a country with a low-medium human development index, among studies that used the MINI as the reference standard (N Studies = 2; N Participants = 245; N major depression = 95)



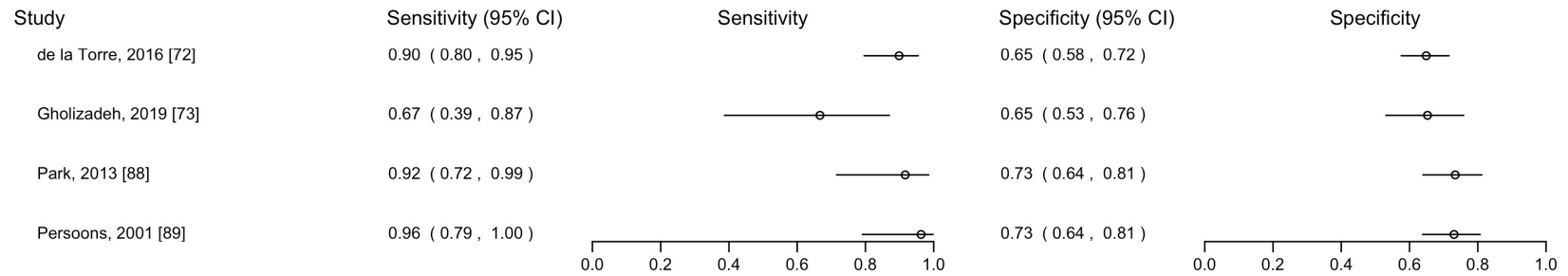
eFigure 2ai. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a non-medical setting, among studies that used the MINI as the reference standard (N Studies = 8; N Participants = 6,792; N major depression = 470)



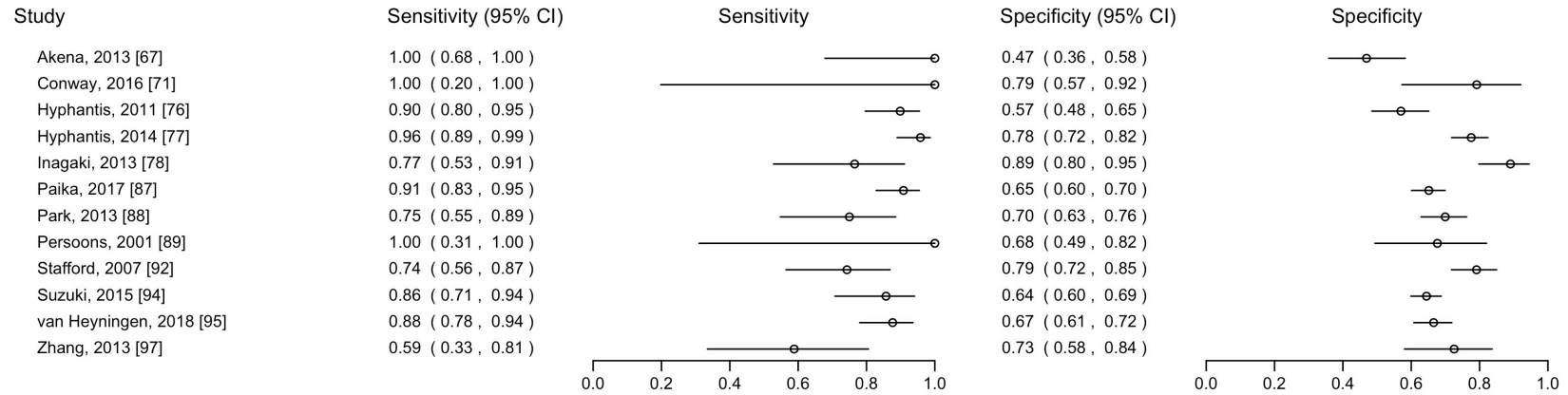
eFigure 2aj. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from a primary care setting, among studies that used the MINI as the reference standard (N Studies = 9; N Participants = 5,092; N major depression = 557)



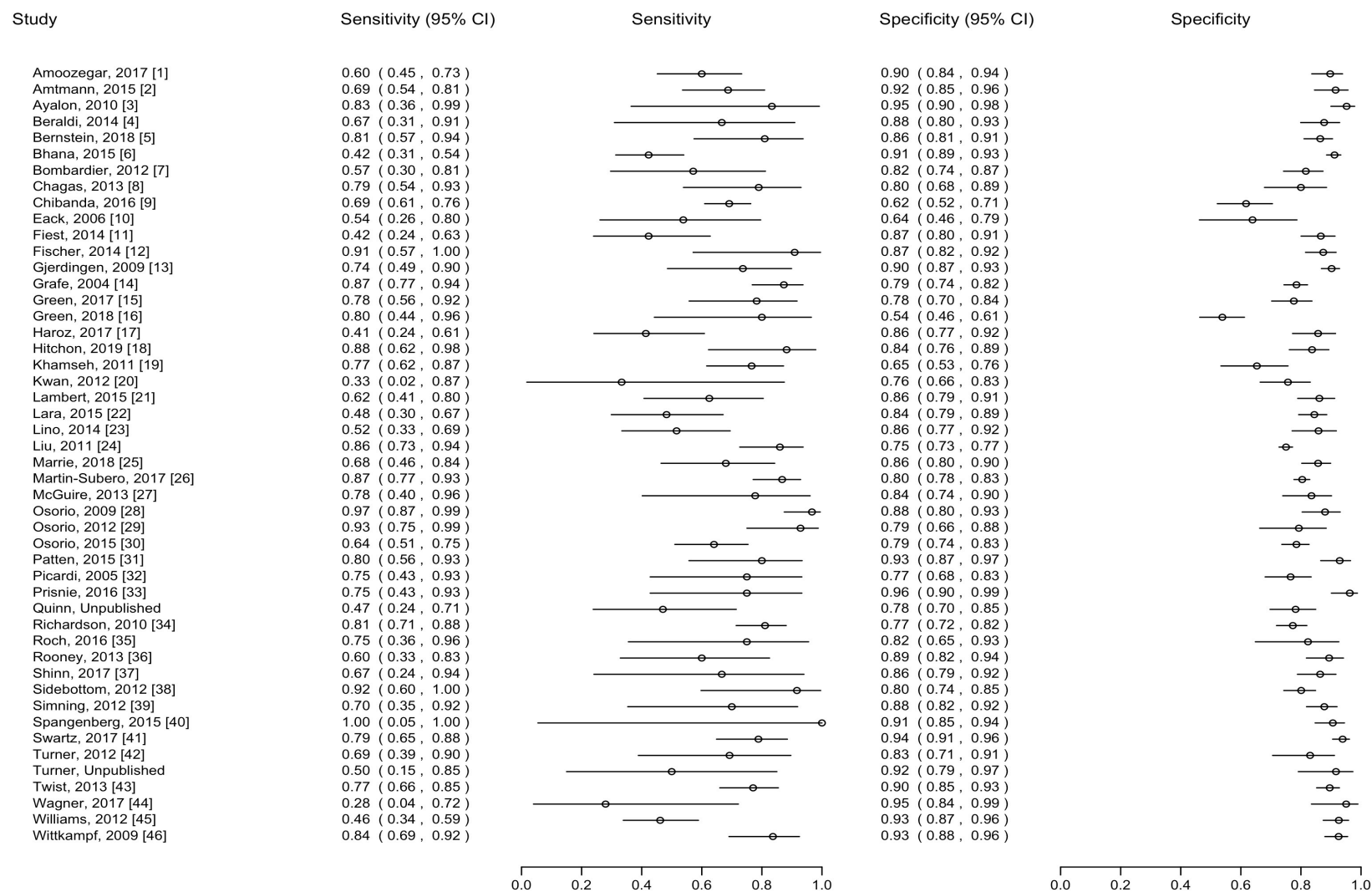
eFigure 2ak. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from an inpatient specialty care setting, among studies that used the MINI as the reference standard (N Studies = 4; N Participants = 619; N major depression = 135)



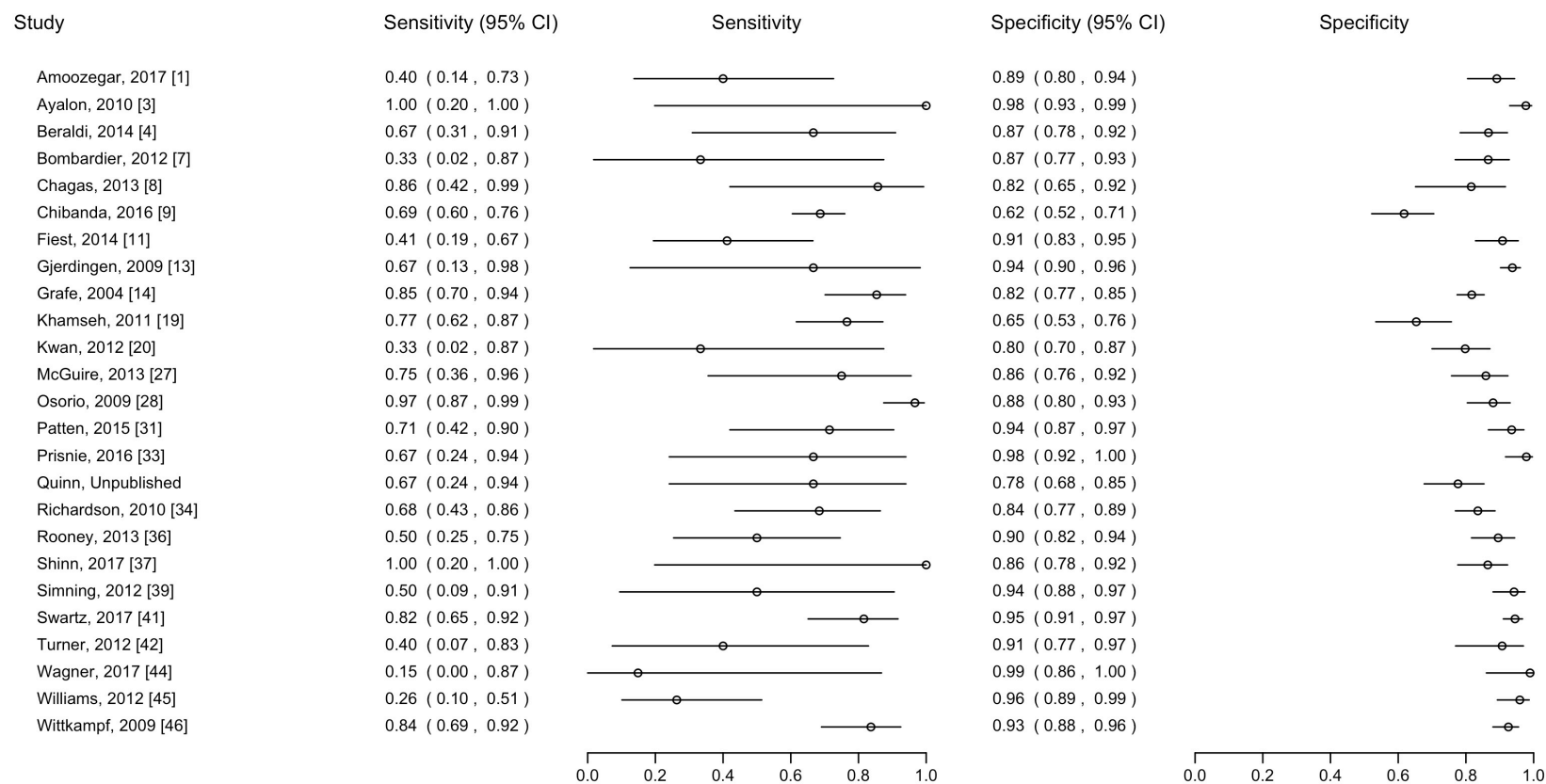
eFigure 2a. Forest plots of sensitivity and specificity estimates for cutoff 2 of the PHQ-2 among participants from an outpatient specialty care setting, among studies that used the MINI as the reference standard (N Studies = 12; N Participants = 2,663; N major depression = 502)



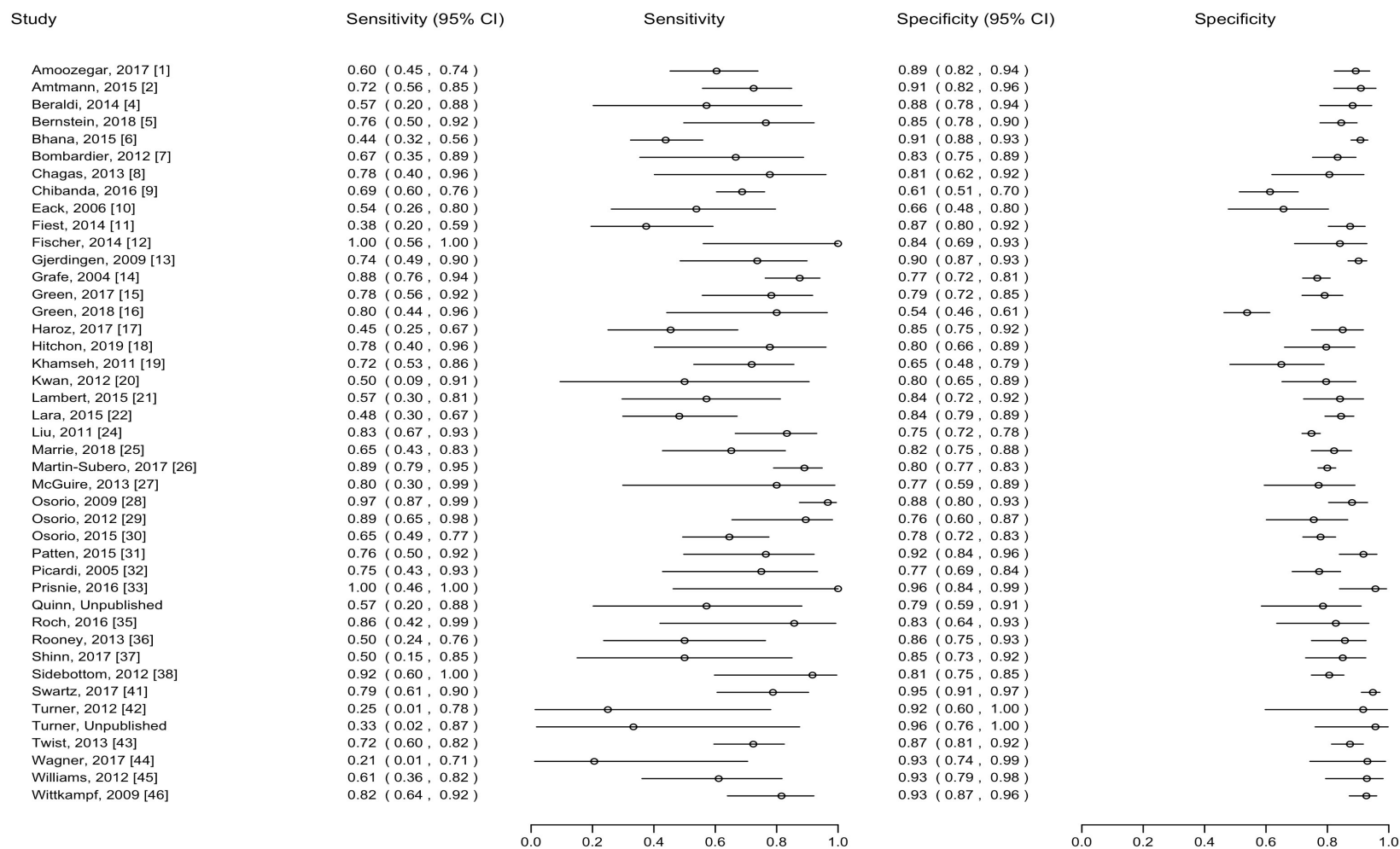
eFigure 2am. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 48; N Participants = 11,703; N major depression = 1,538)



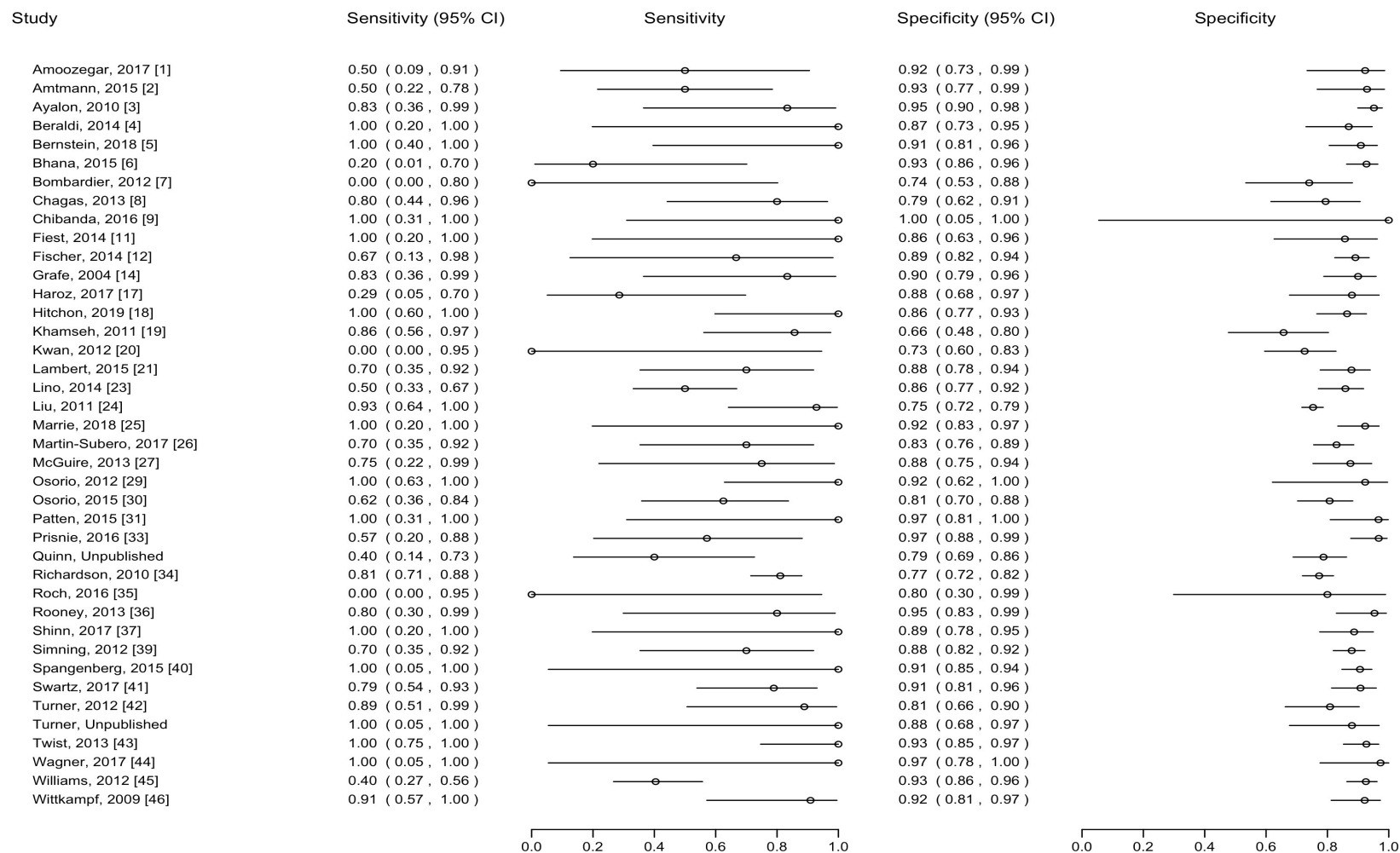
eFigure 2an. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants verified to not currently be diagnosed or receiving treatment for a mental health problem, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 25; N Participants = 3,708; N major depression = 527)



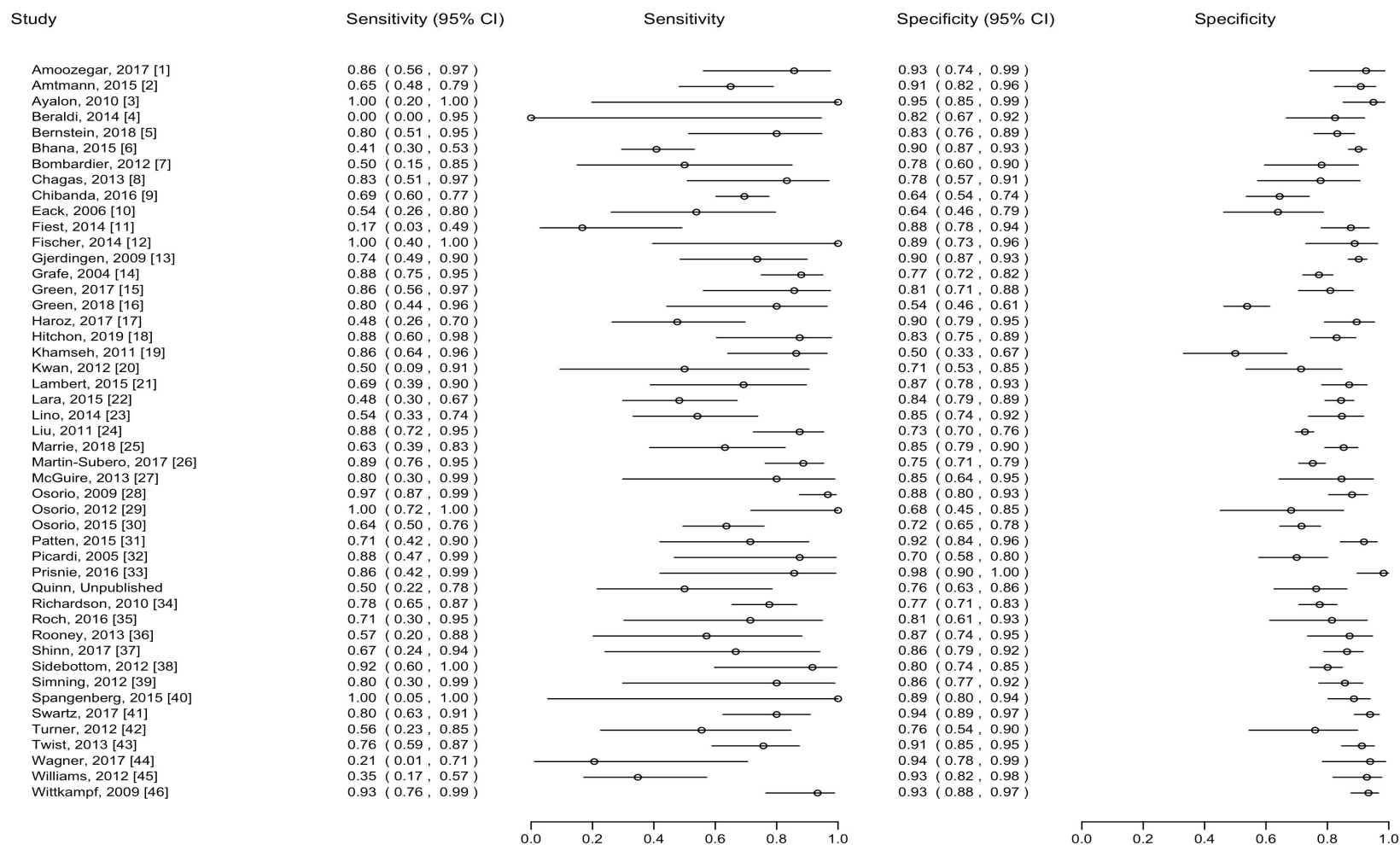
eFigure 2ao. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants aged < 60, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 43; N Participants = 7,759; N major depression = 1,117)



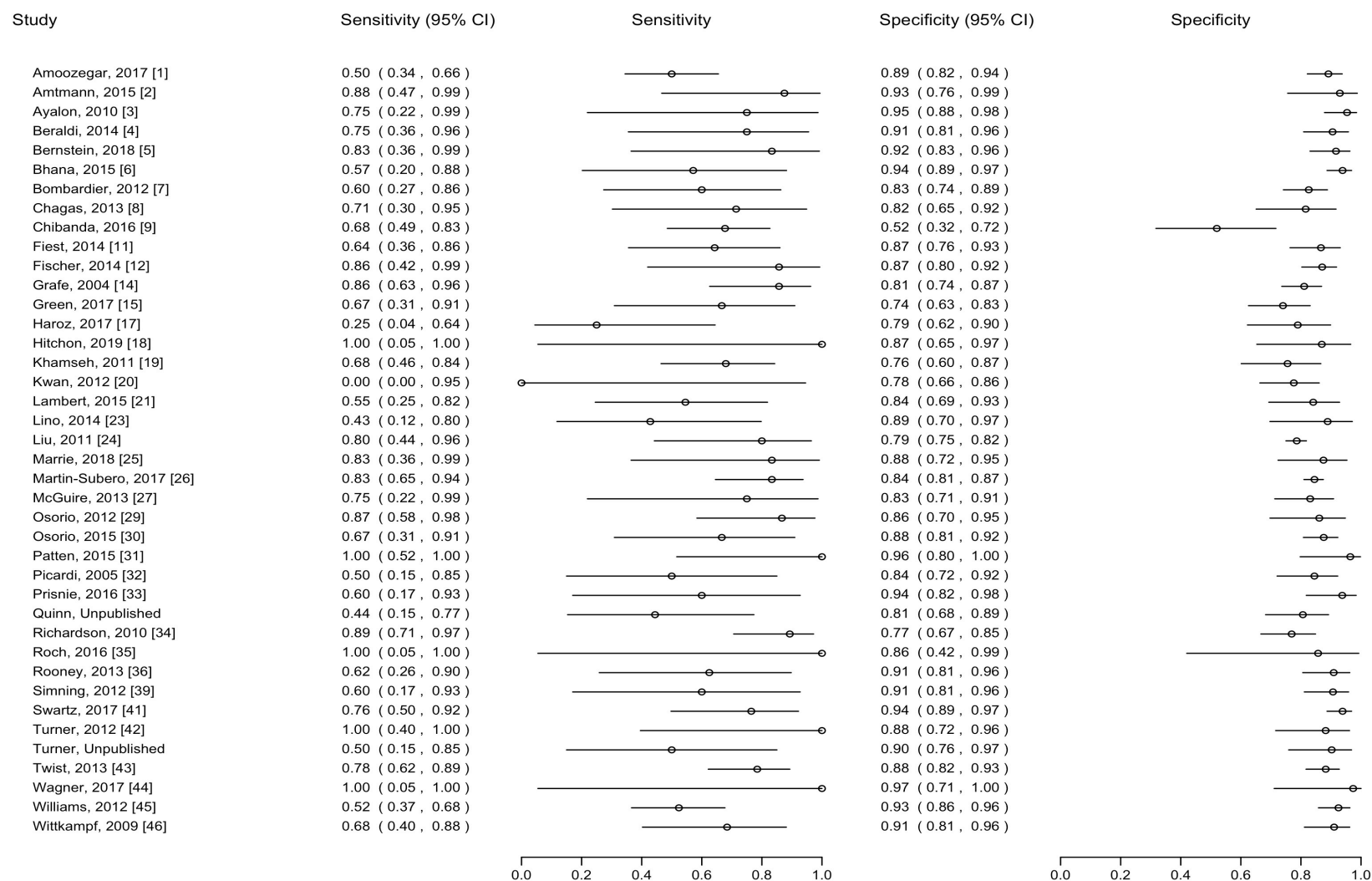
eFigure 2ap. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants aged ≥ 60 , among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 40; N Participants = 3,875; N major depression = 415)



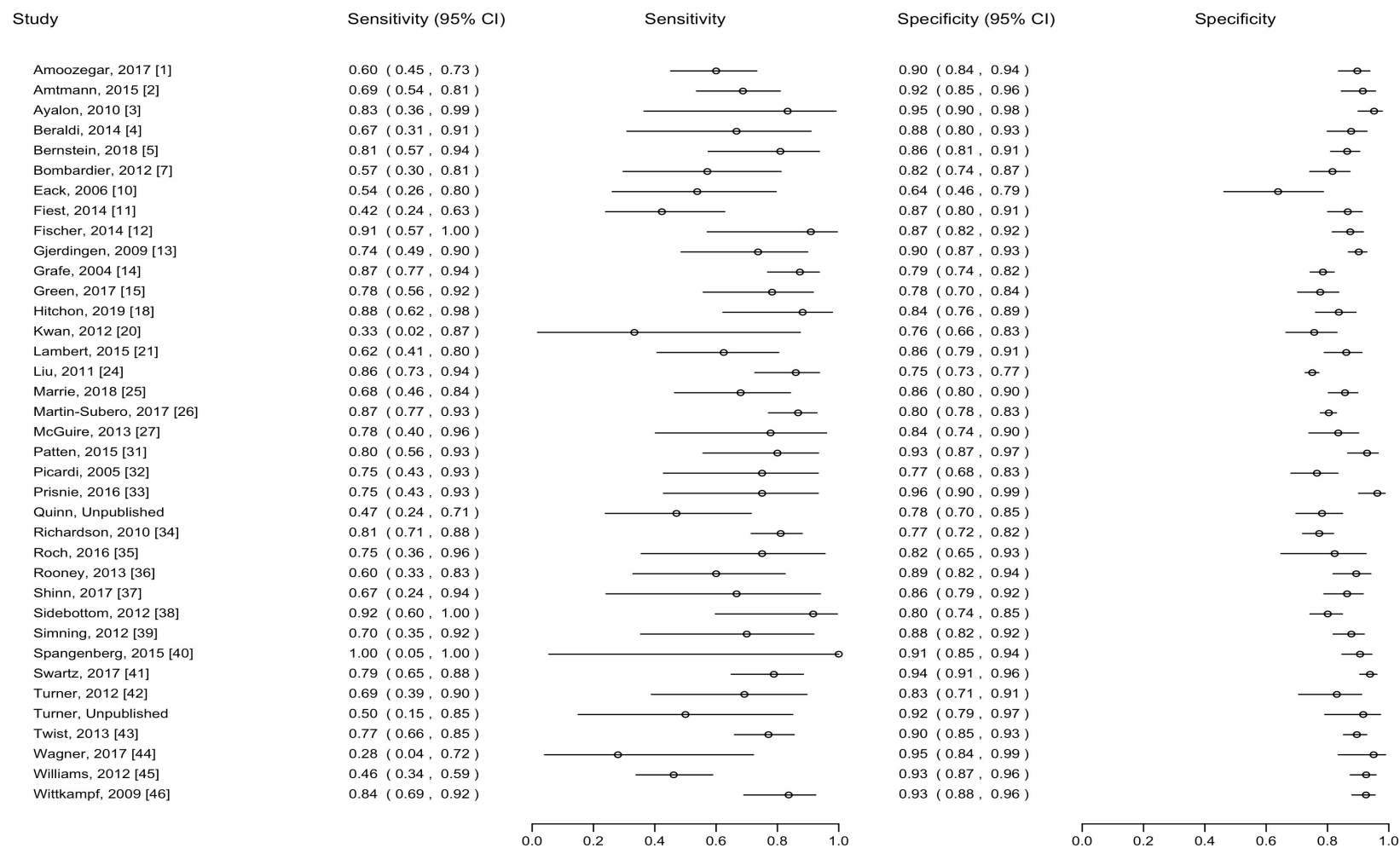
eFigure 2aq. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among women, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 47; N Participants = 7,280; N major depression = 1,054)



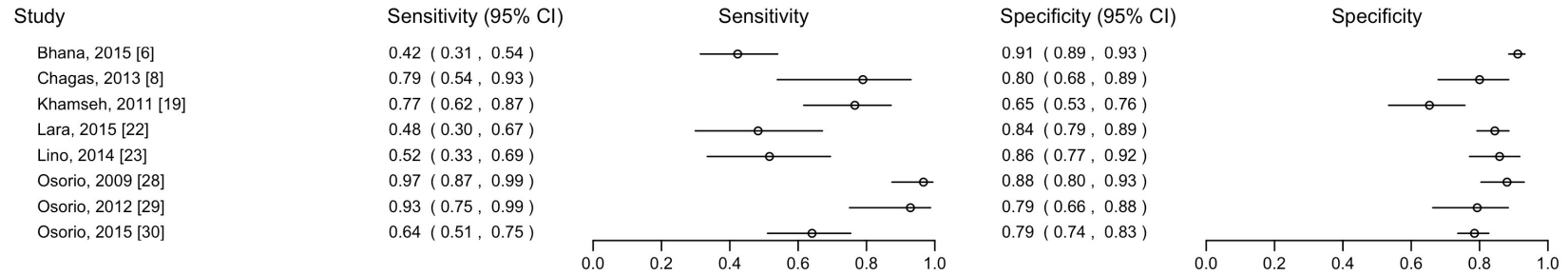
eFigure 2ar. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among men, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 40; N Participants = 4,345; N major depression = 484)



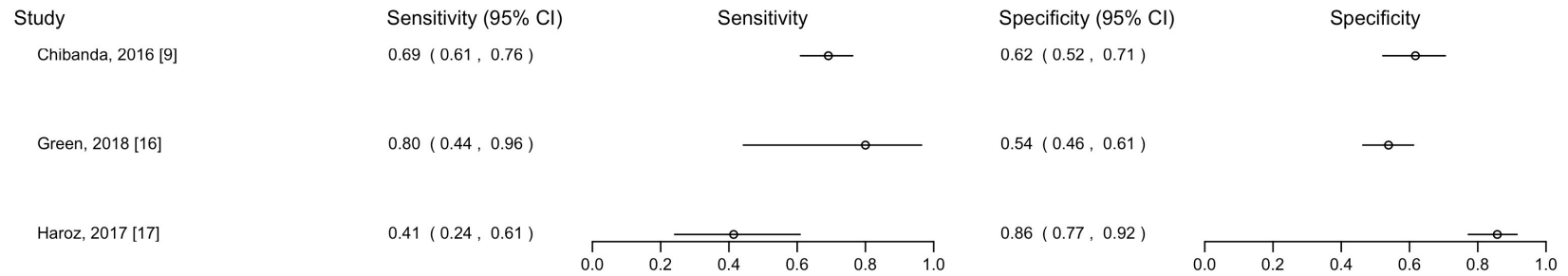
eFigure 2as. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a country with a very high human development index, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 37; N Participants = 9,156; N major depression = 994)



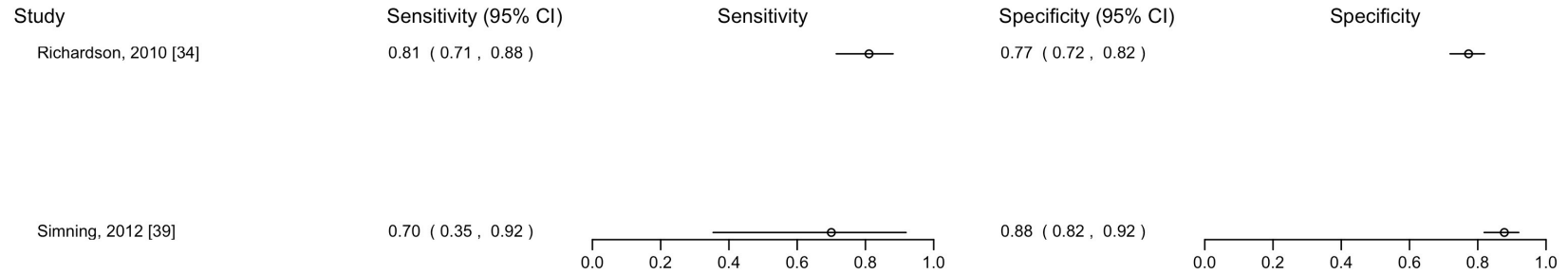
eFigure 2at. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a country with a high human development index, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 8; N Participants = 1,957; N major depression = 356)



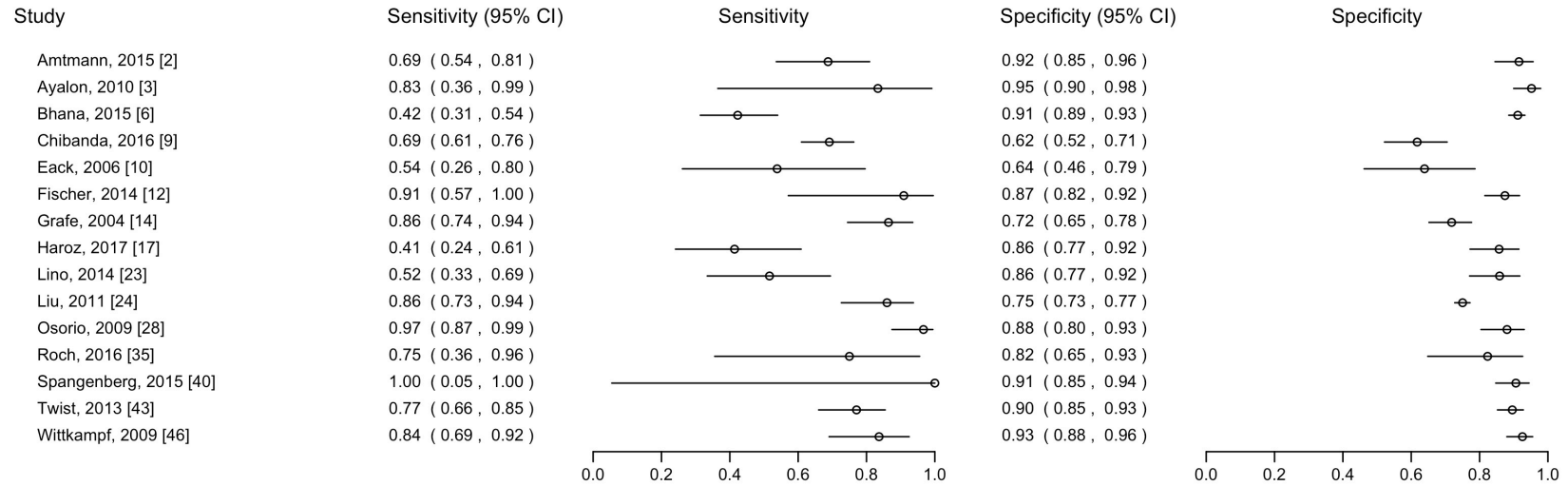
eFigure 2au. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a country with a low-medium human development index, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 3; N Participants = 590; N major depression = 188)



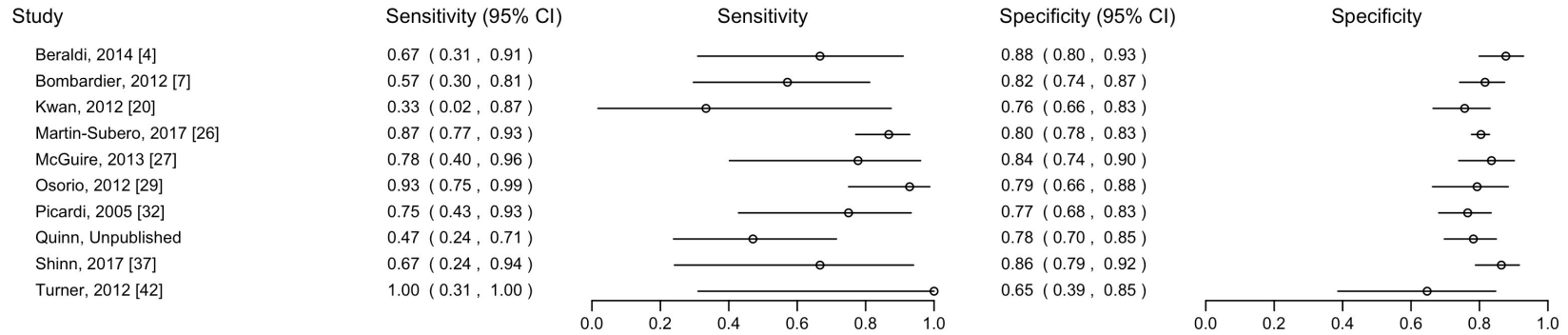
eFigure 2av. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a non-medical setting, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 2; N Participants = 567; N major depression = 105)



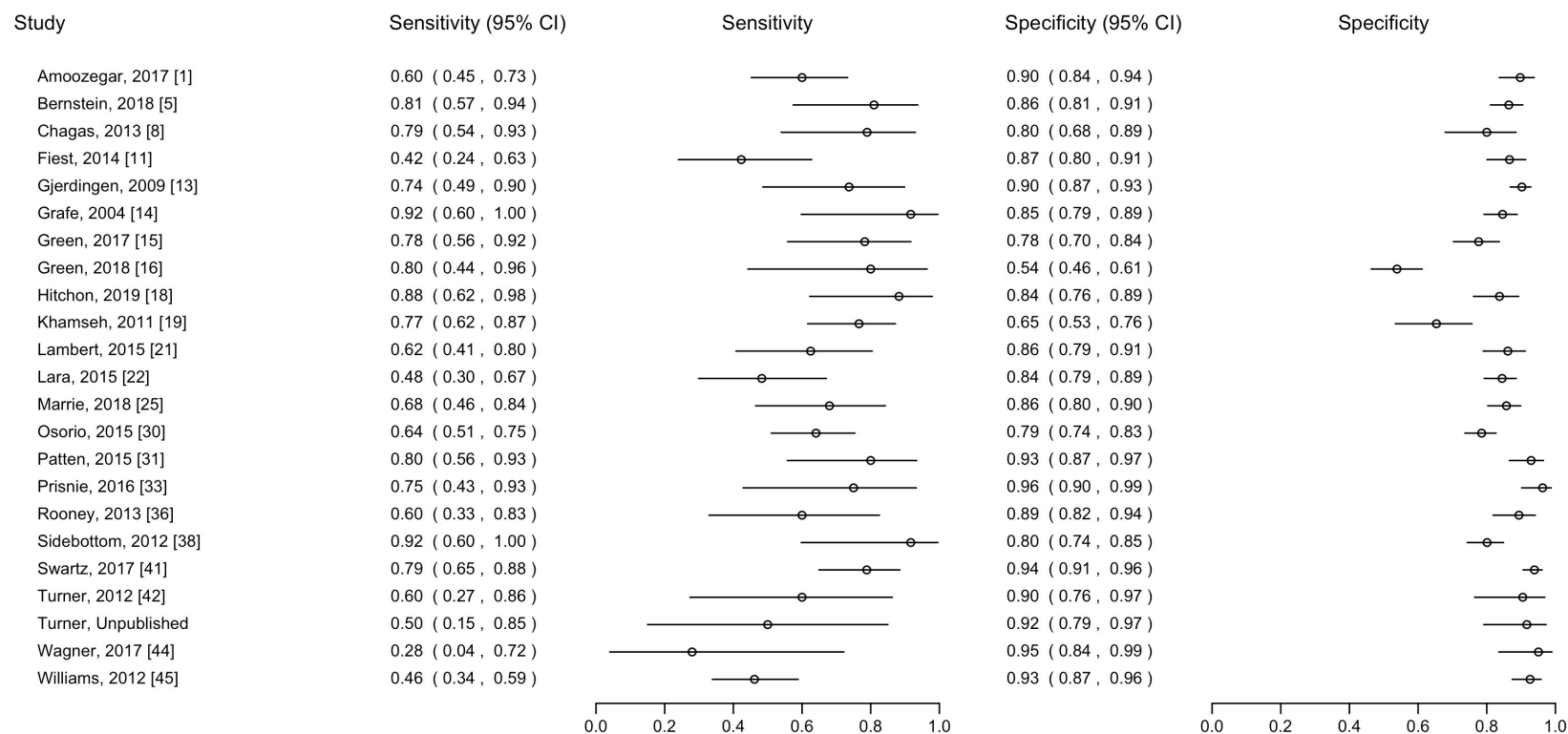
eFigure 2aw. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a primary care setting, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 15; N Participants = 4,569; N major depression = 667)



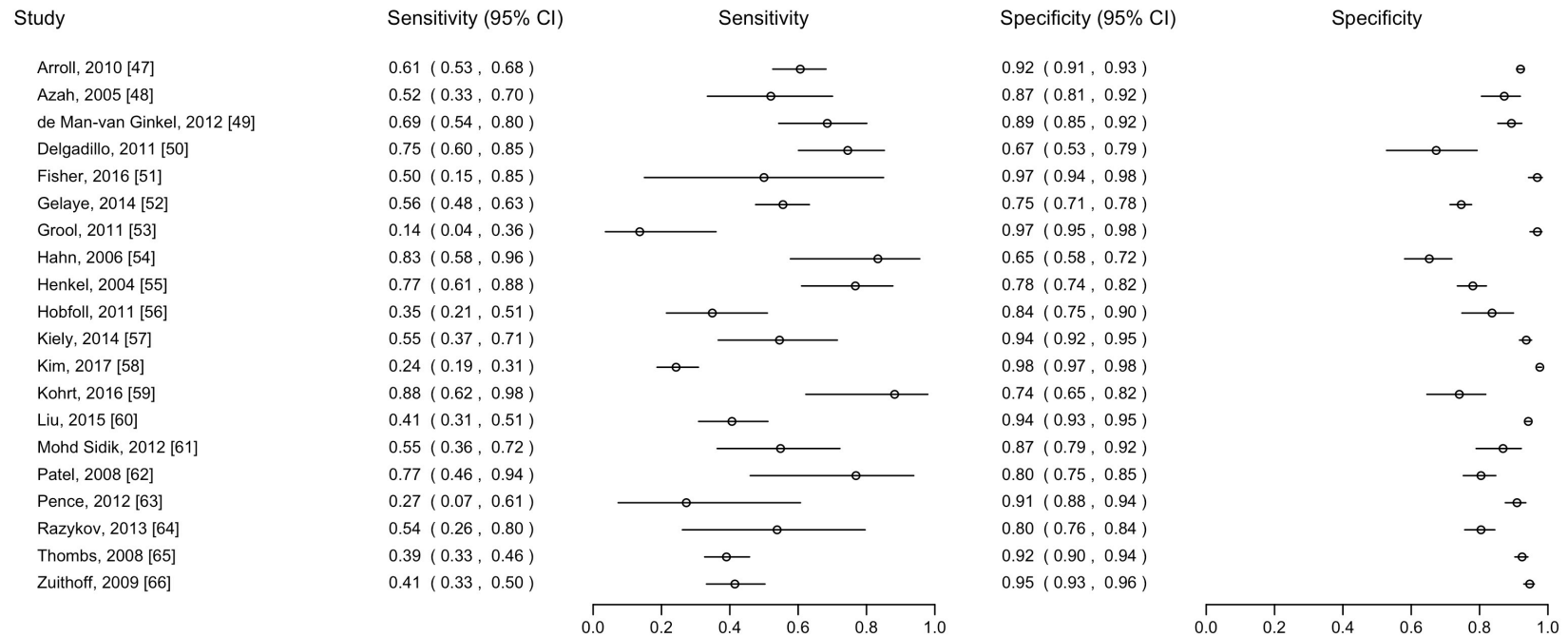
eFigure 2ax. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from an inpatient specialty care setting, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 10; N Participants = 2,019; N major depression = 184)



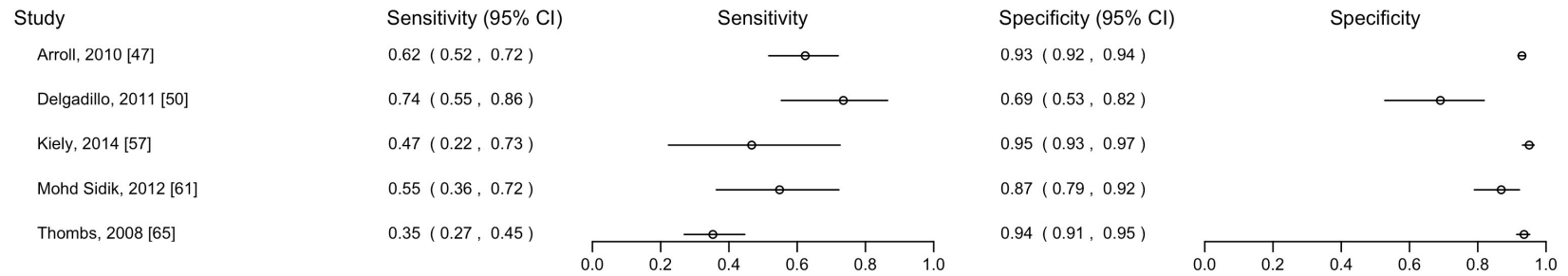
eFigure 2ay. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from an outpatient specialty care setting, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 23; N Participants = 4,548; N major depression = 582)



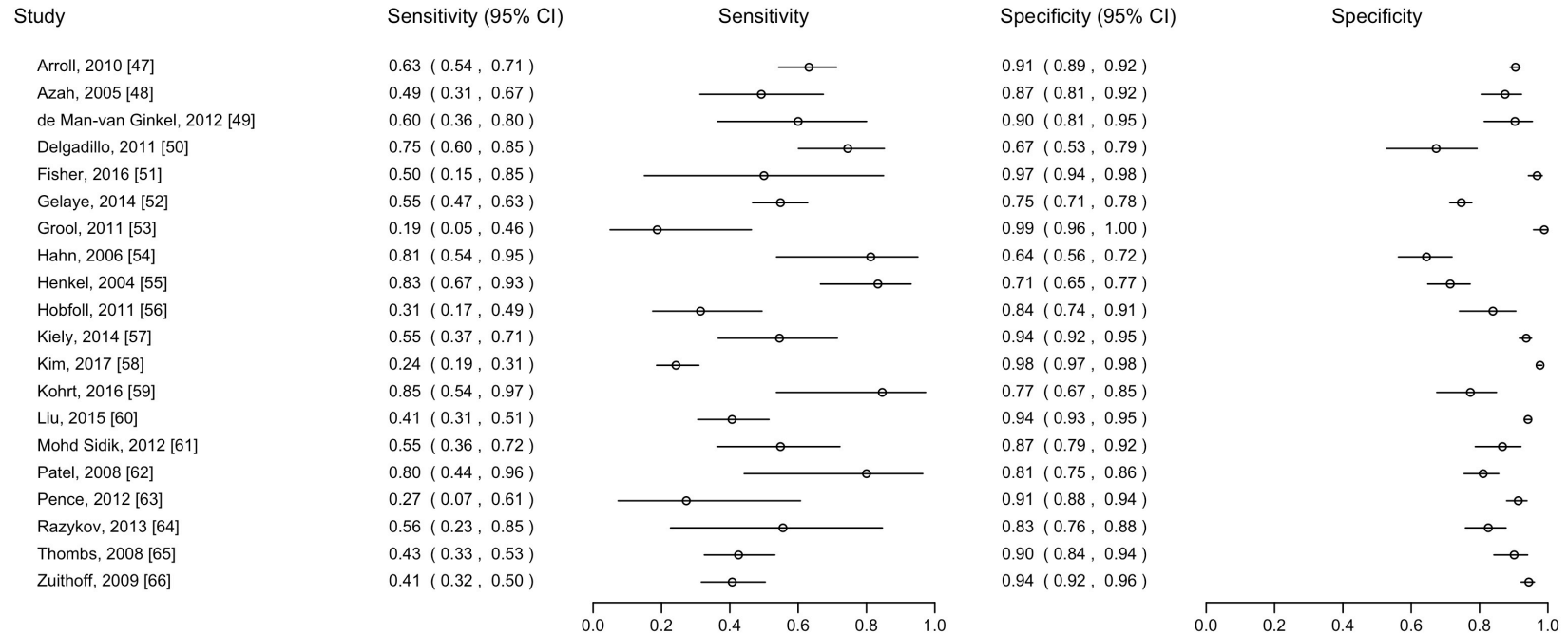
eFigure 2az. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 20; N Participants = 17,319; N major depression = 1,365)



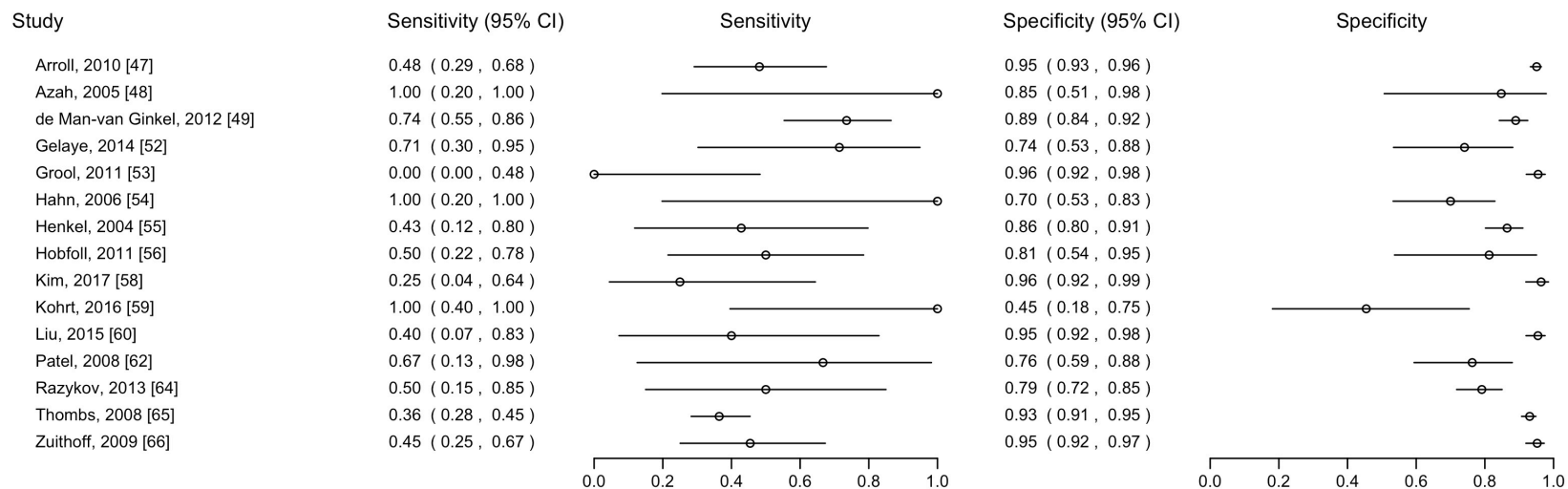
eFigure 2aaa. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants verified to not currently be diagnosed or receiving treatment for a mental health problem, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 5; N Participants = 4,050; N major depression = 292)



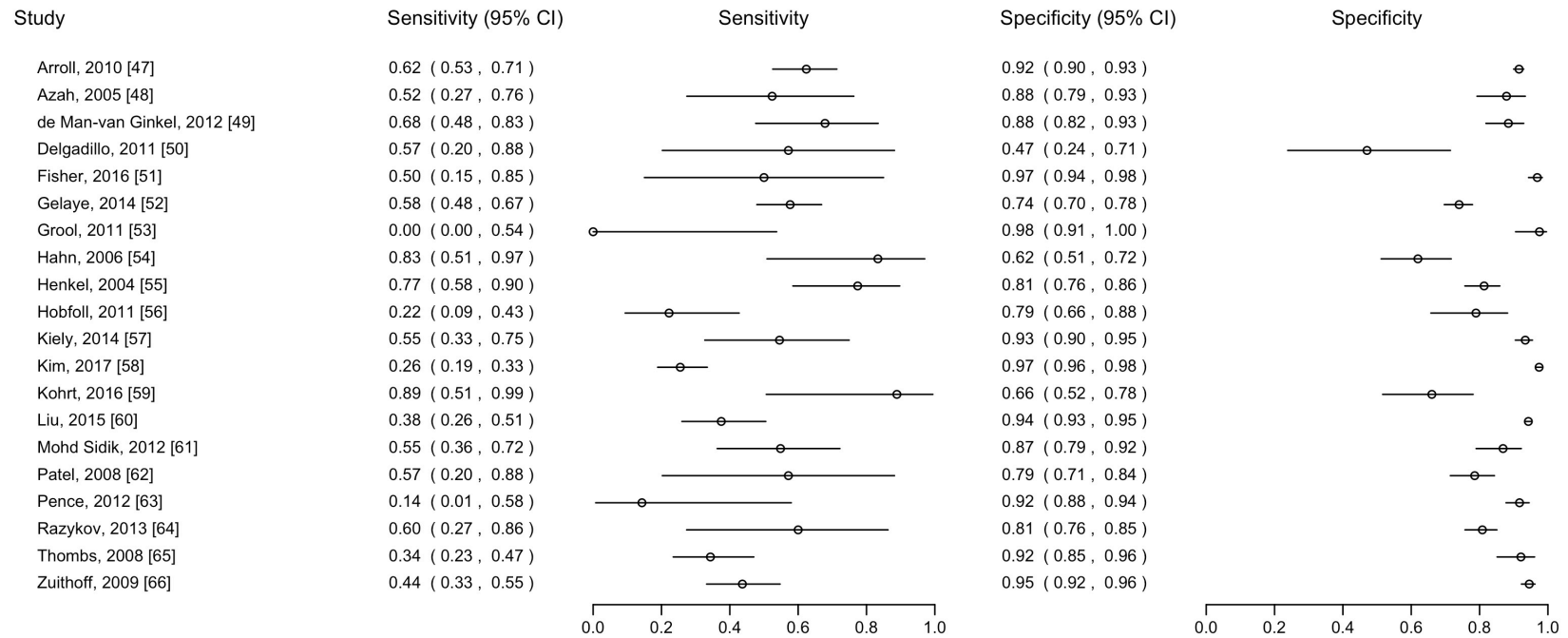
eFigure 2aab. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants aged < 60, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 20; N Participants = 13,901; N major depression = 1,097)



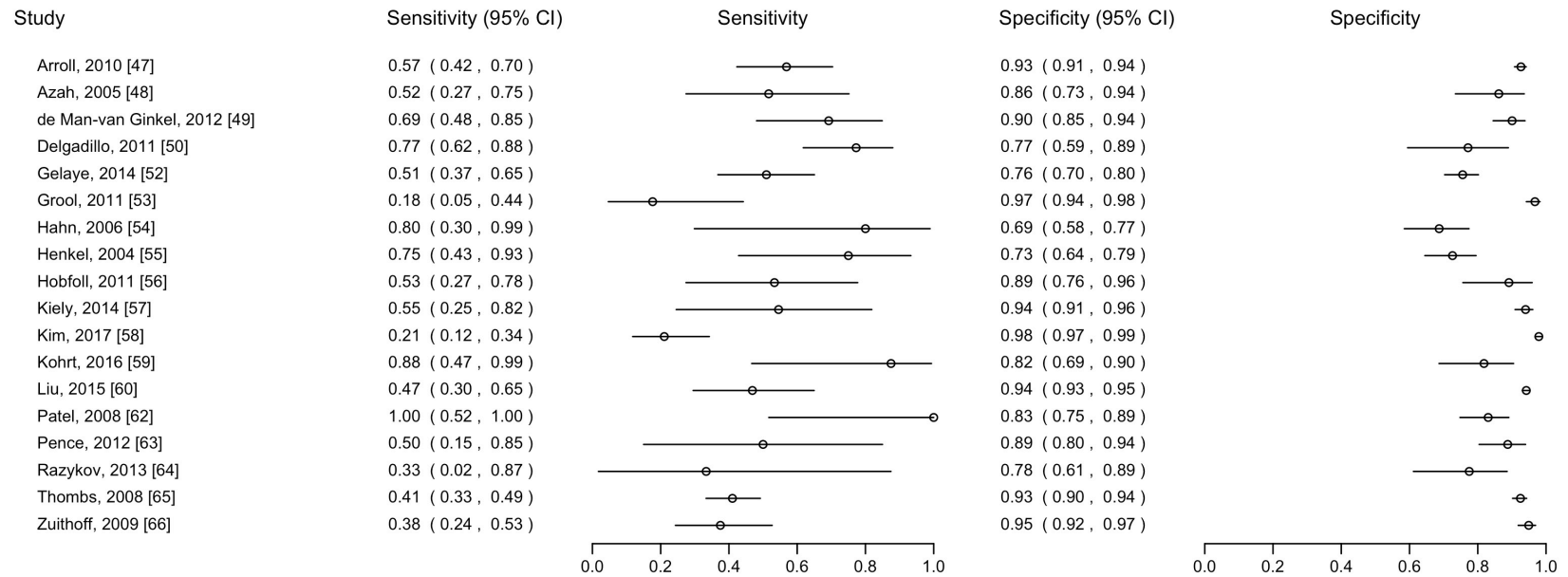
eFigure 2aac. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants aged ≥ 60 , among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 15; N Participants = 3,400; N major depression = 268)



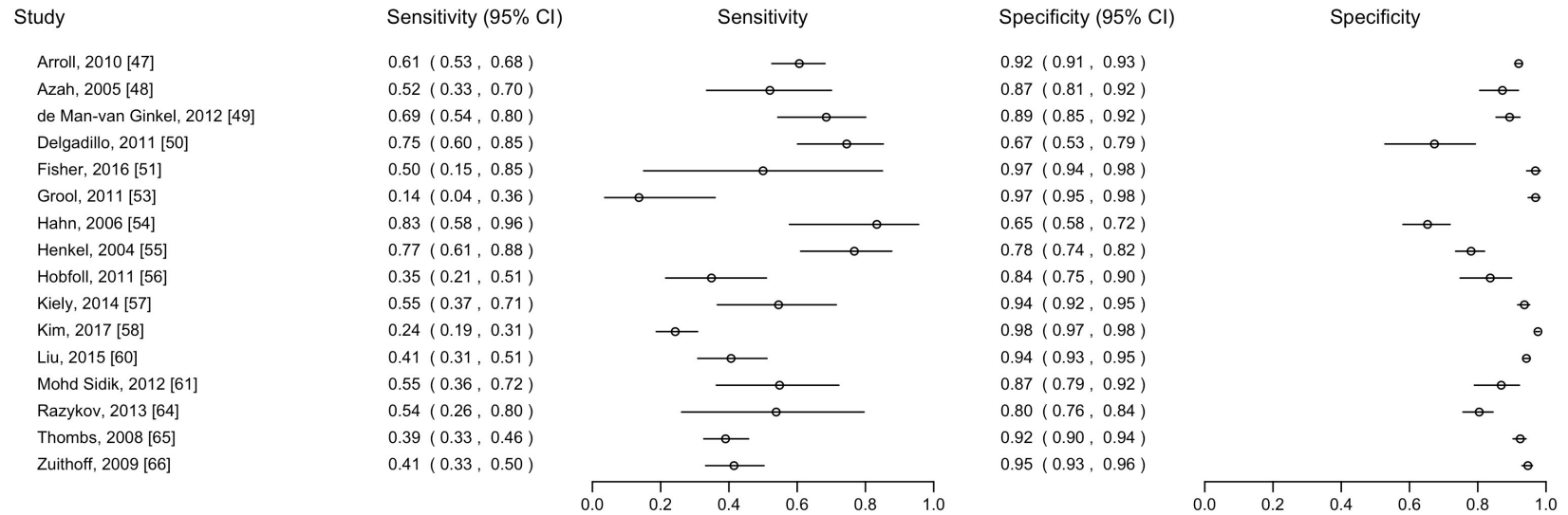
eFigure 2aad. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among women, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 20; N Participants = 9,690; N major depression = 802)



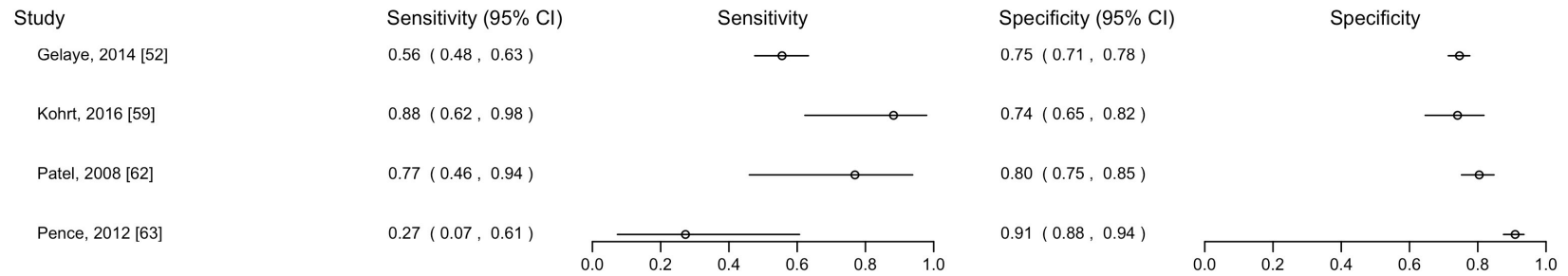
eFigure 2aae. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among men, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 18; N Participants = 7,619; N major depression = 561)



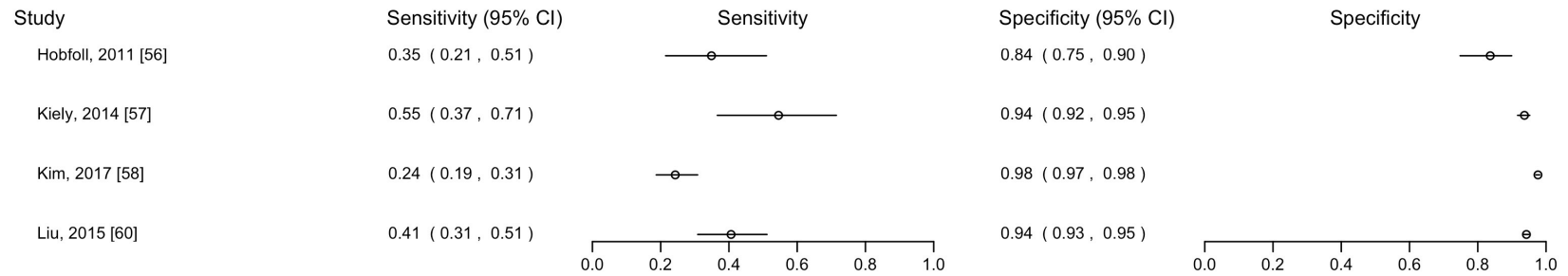
eFigure 2aaf. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a country with a very high human development index, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 16; N Participants = 15,574; N major depression = 1,162)



eFigure 2aag. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a country with a low-medium human development index, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 4; N Participants = 1,745; N major depression = 203)

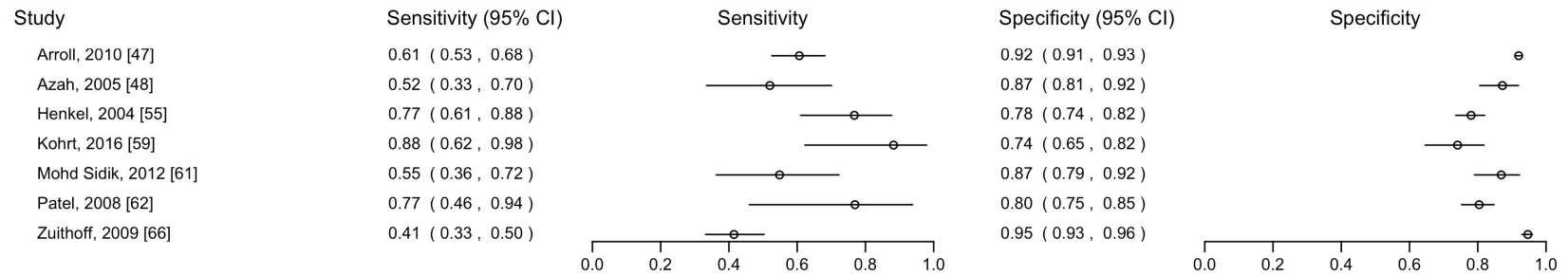


eFigure 2aah. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a non-medical setting, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 4; N Participants = 8,316; N major depression = 378)

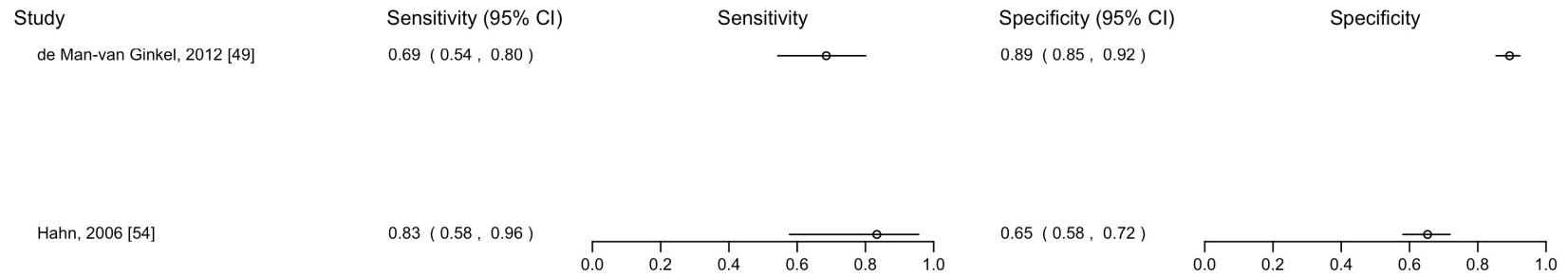


eFigure 2aai. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a primary care setting, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 7; N

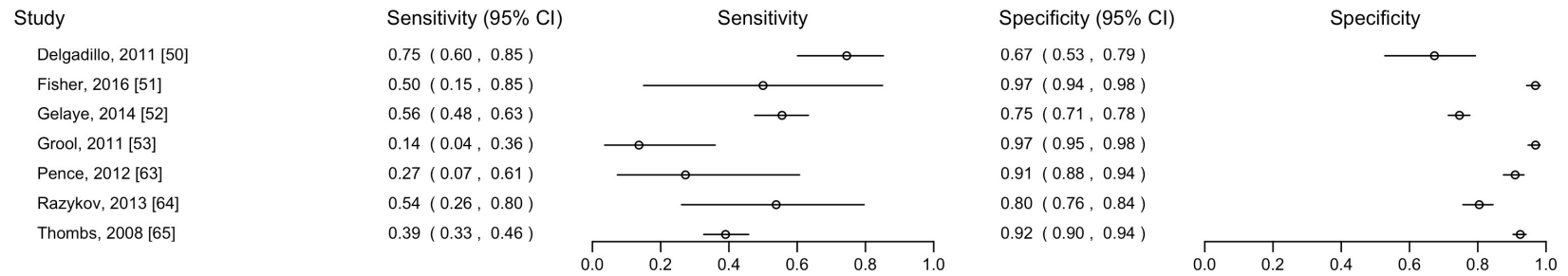
Participants = 4,789; N major depression = 429)



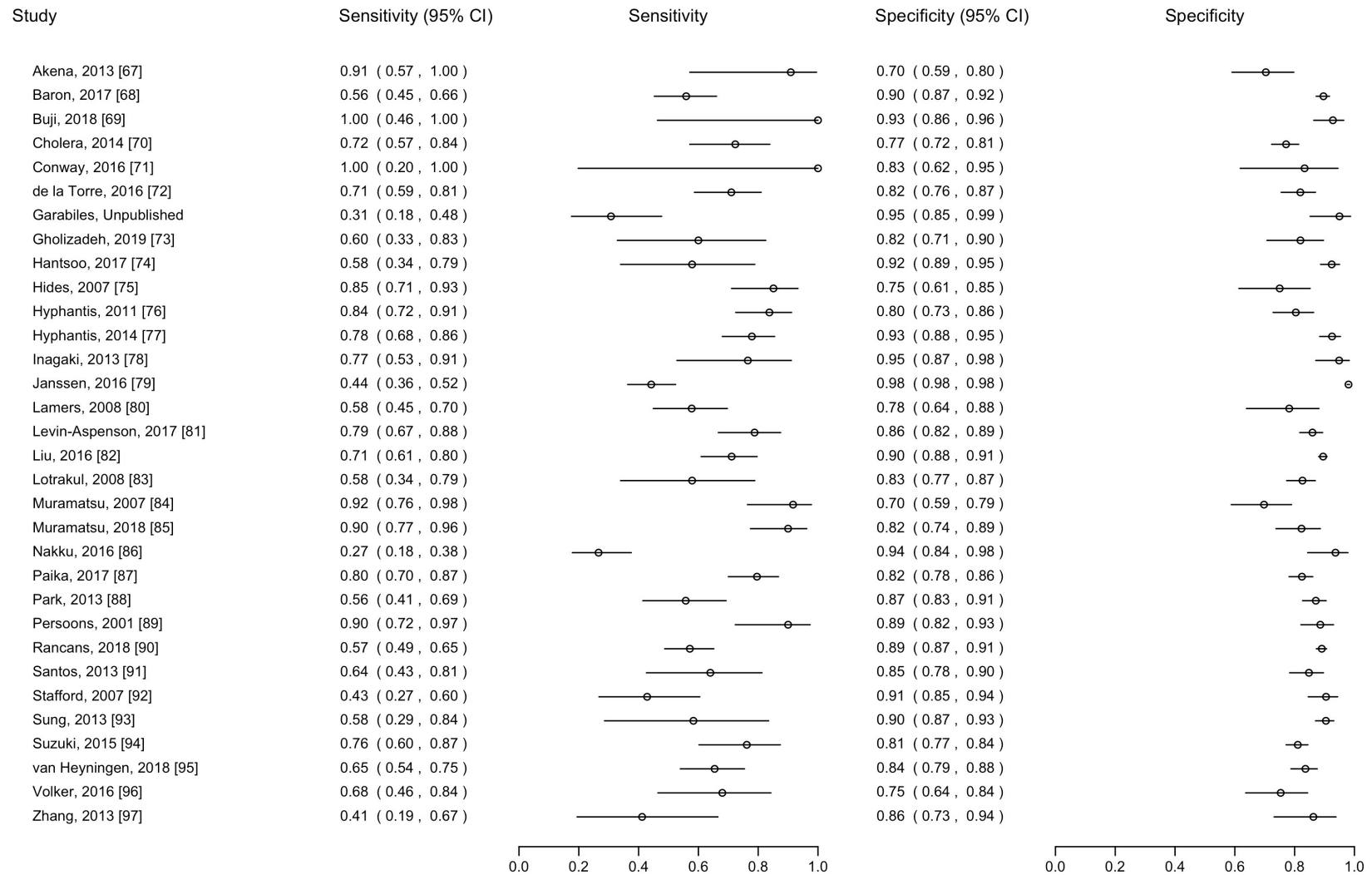
eFigure 2aaj. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from an inpatient specialty care setting, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 2; N Participants = 593; N major depression = 72)



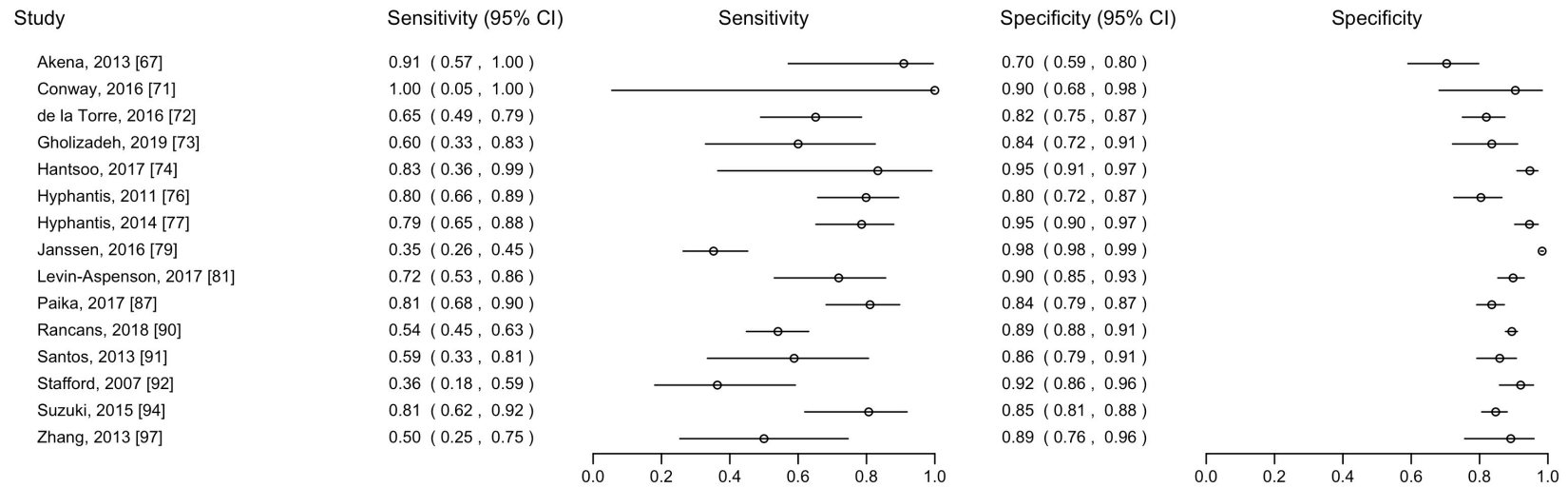
eFigure 2aak. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from an outpatient specialty care setting, among studies that used a fully structured diagnostic interview as the reference standard (N Studies = 7; N Participants = 3,621; N major depression = 486)



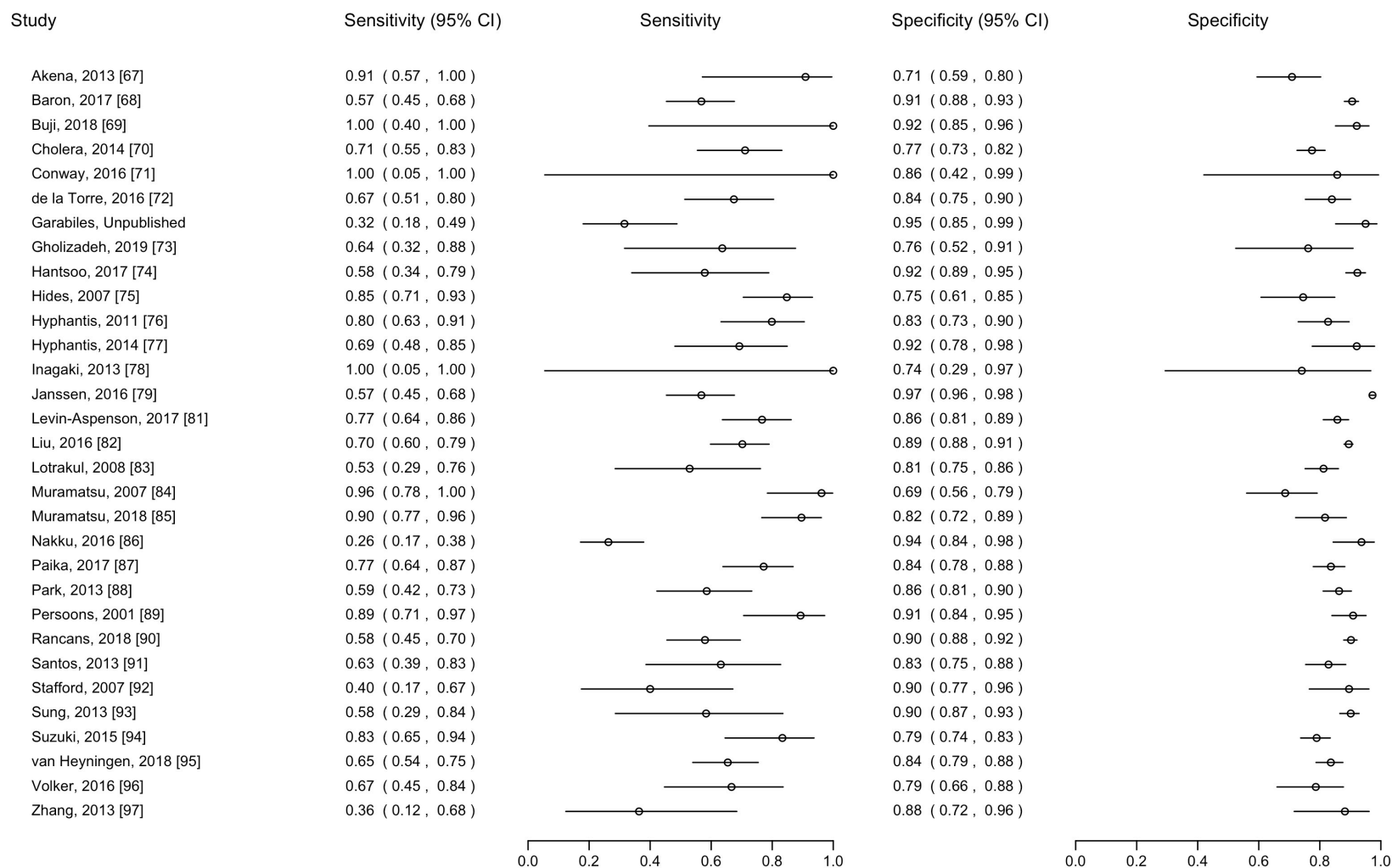
eFigure 2aal. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2, among studies that used the MINI as the reference standard (N Studies = 32; N Participants = 15,296; N major depression = 1,669)



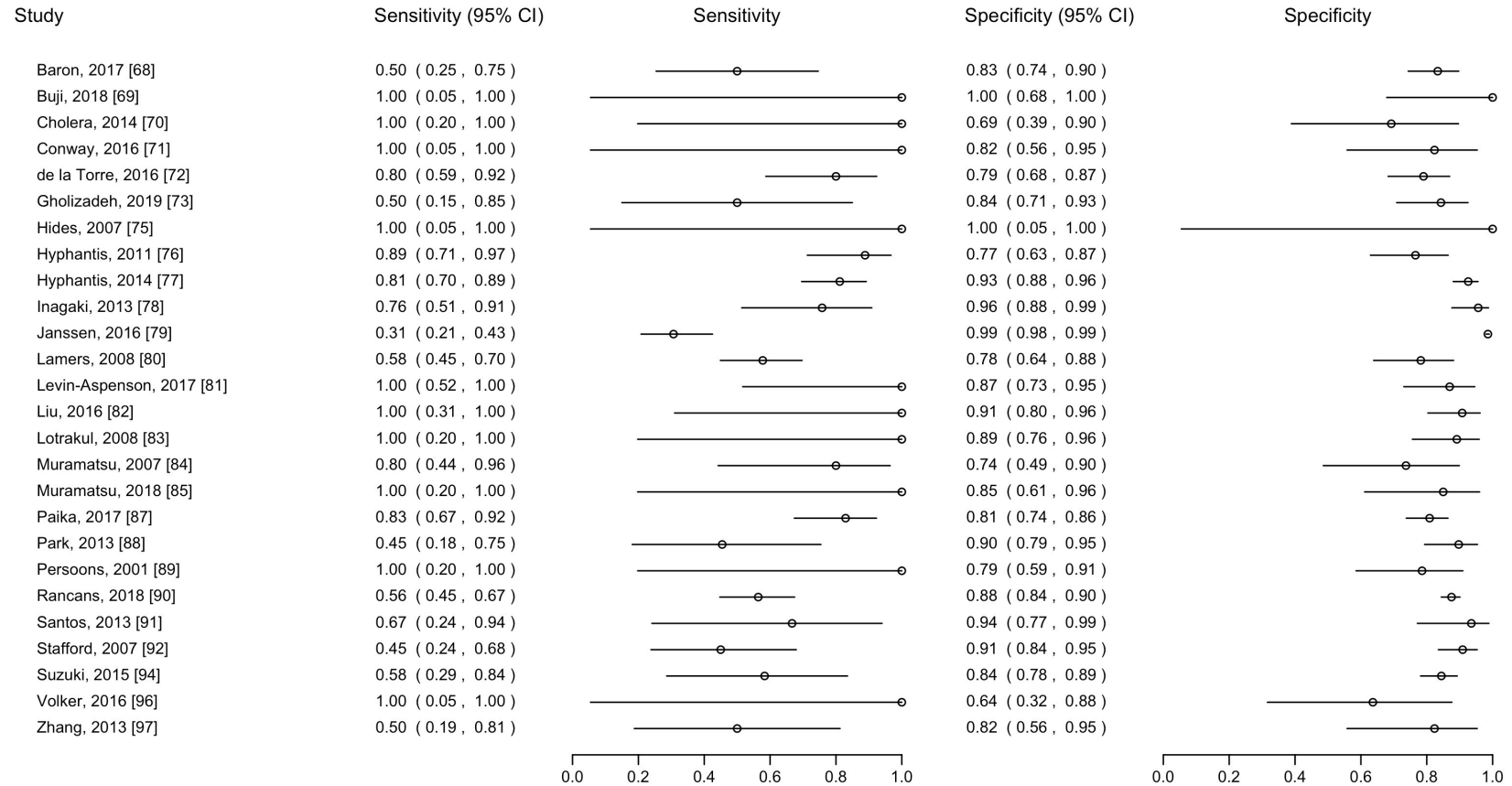
eFigure 2aam. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants verified to not currently be diagnosed or receiving treatment for a mental health problem, among studies that used the MINI as the reference standard (N Studies = 15; N Participants = 8,390; N major depression = 581)



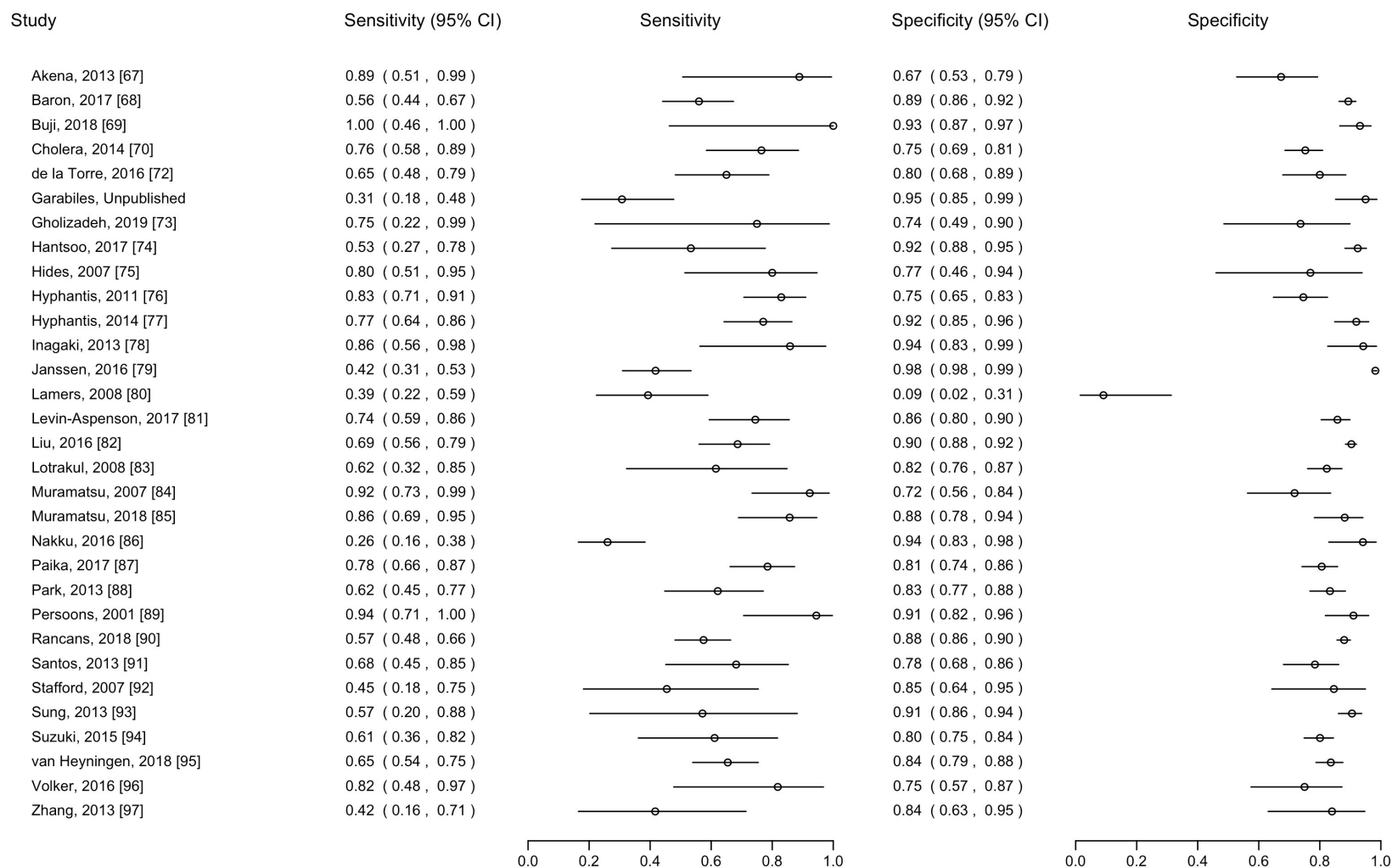
eFigure 2aan. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants aged < 60, among studies that used the MINI as the reference standard (N Studies = 31; N Participants = 10,071; N major depression = 1,153)



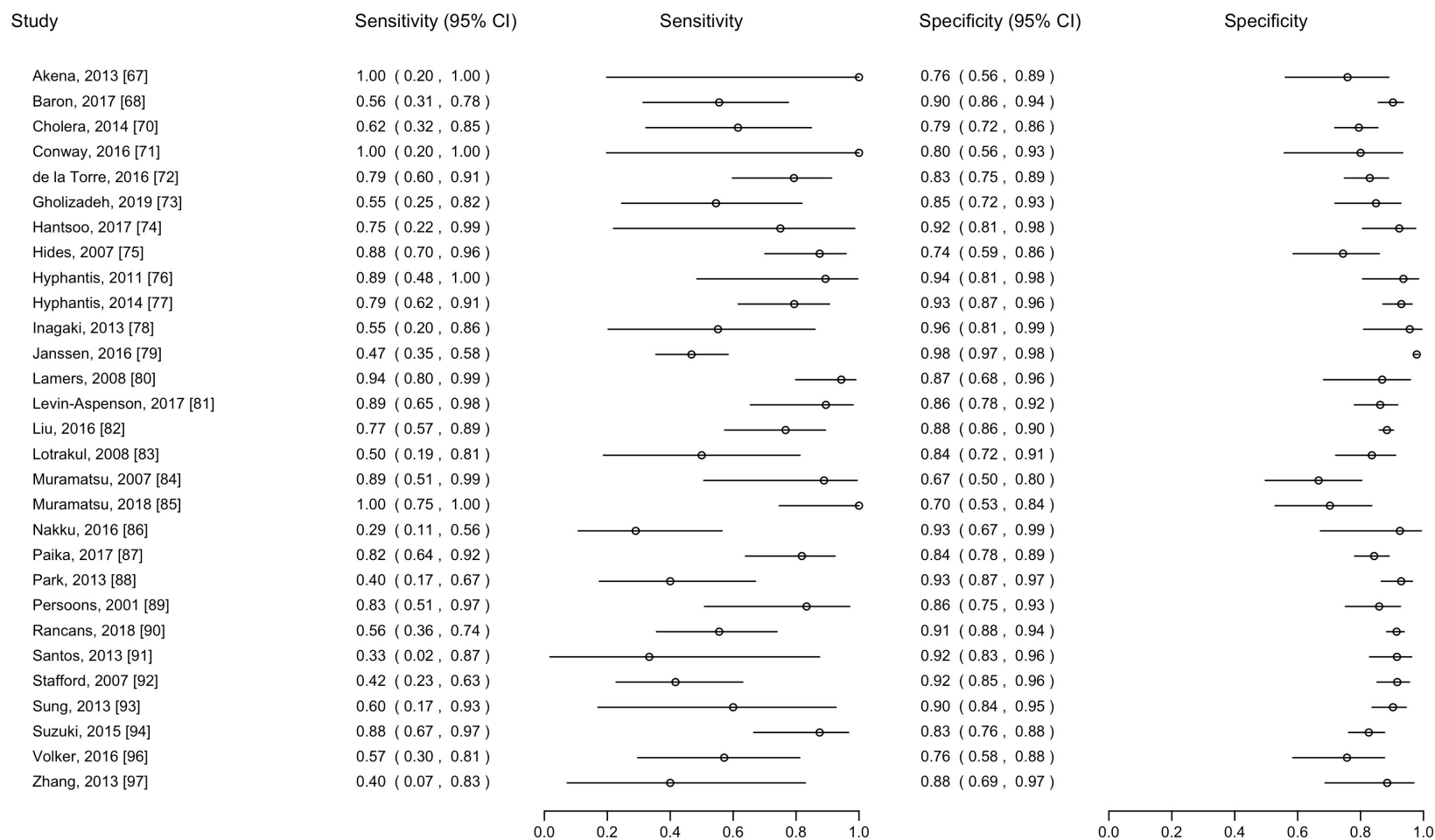
eFigure 2aao. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants aged ≥ 60 , among studies that used the MINI as the reference standard (N Studies = 26; N Participants = 5,192; N major depression = 506)



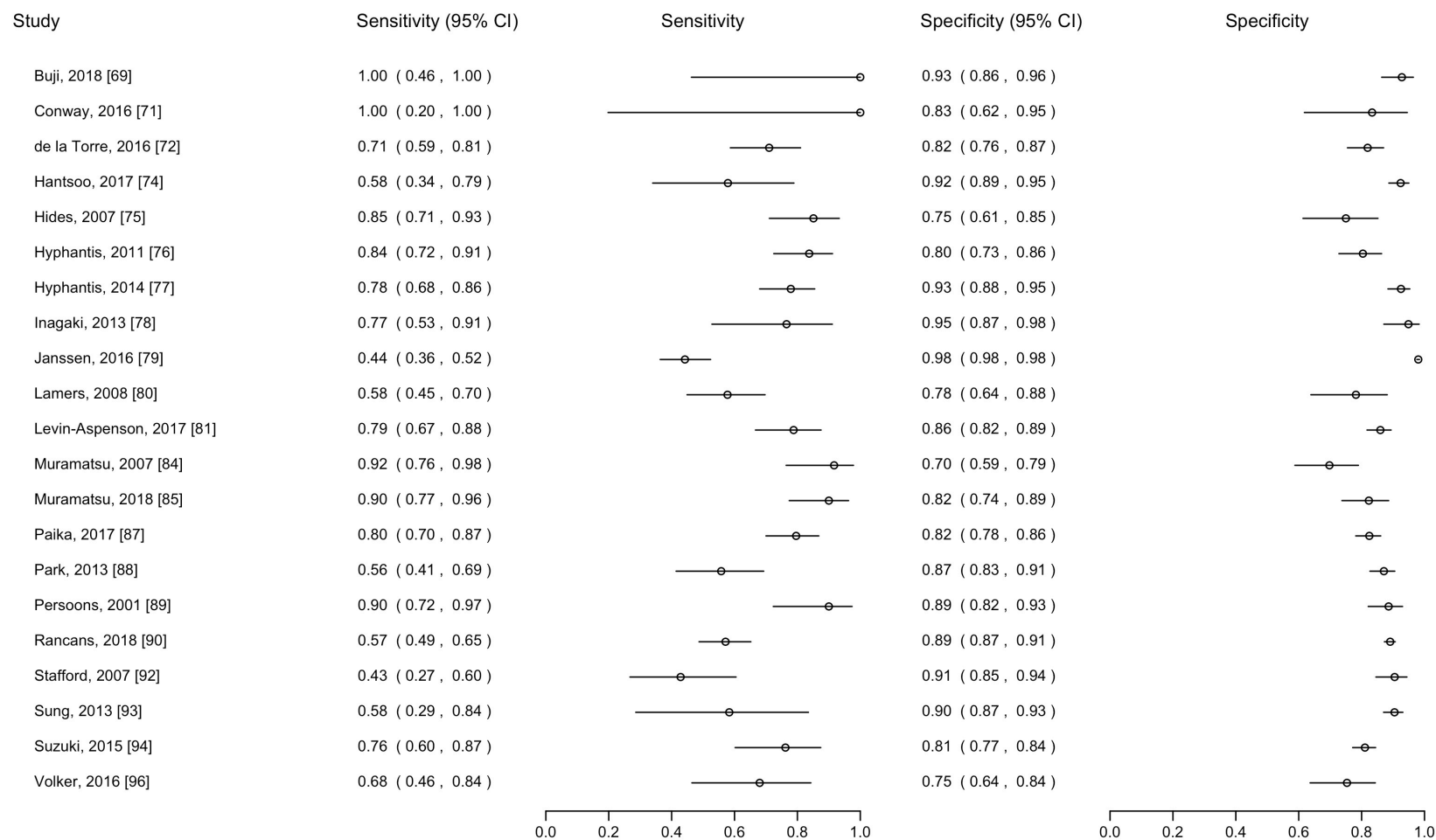
eFigure 2aap. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among women, among studies that used the MINI as the reference standard (N Studies = 31; N Participants = 9,053; N major depression = 1,138)



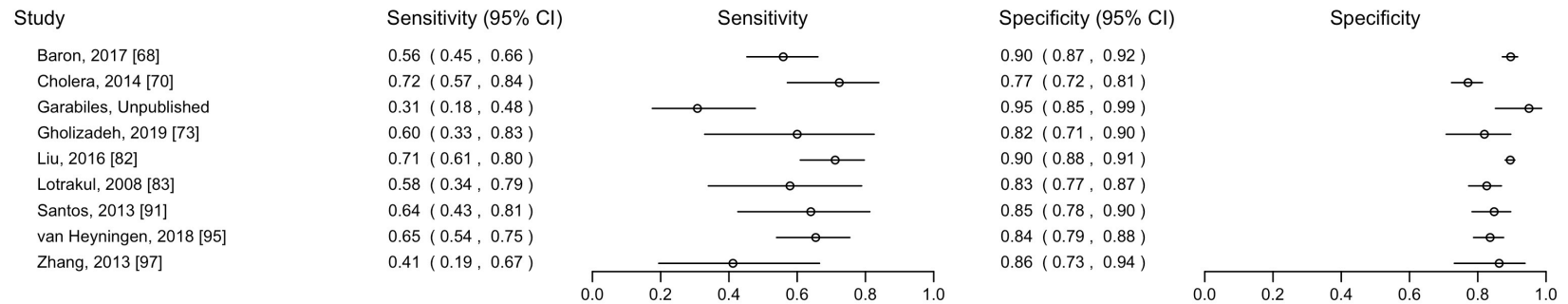
eFigure 2aaq. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among men, among studies that used the MINI as the reference standard (N Studies = 29; N Participants = 6,225; N major depression = 530)



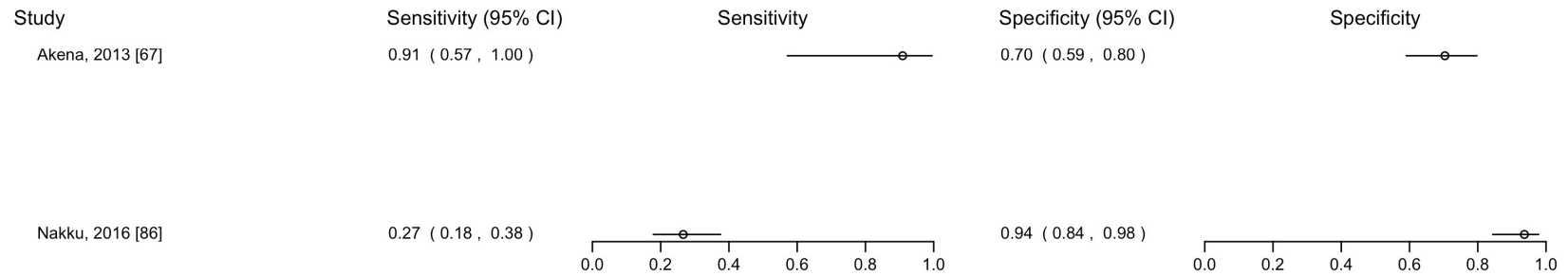
eFigure 2aar. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a country with a very high human development index, among studies that used the MINI as the reference standard (N Studies = 21; N Participants = 10,699; N major depression = 1,141)



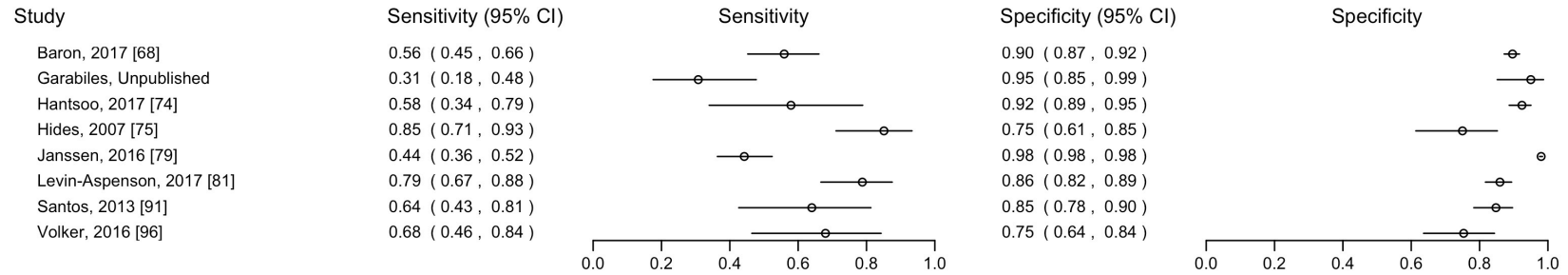
eFigure 2aas. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a country with a high human development index, among studies that used the MINI as the reference standard (N Studies = 9; N Participants = 4,352; N major depression = 433)



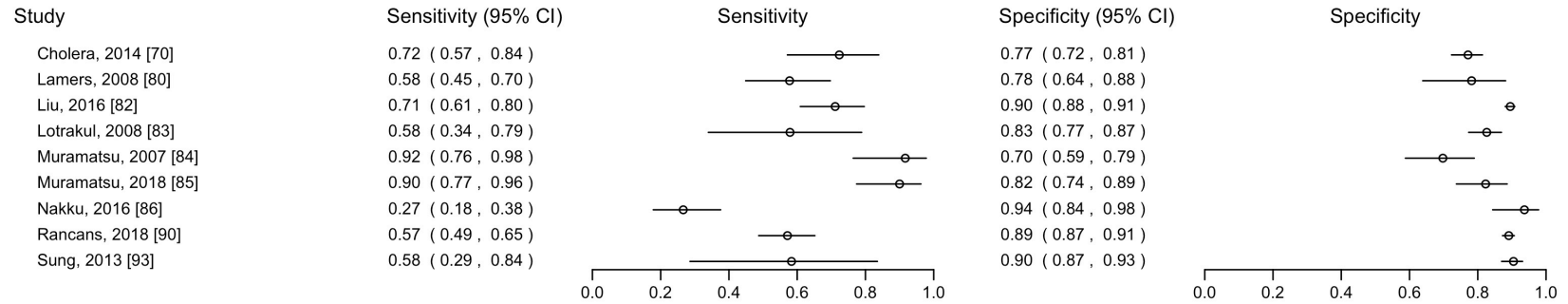
eFigure 2aat. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a country with a low-medium human development index, among studies that used the MINI as the reference standard (N Studies = 2; N Participants = 245; N major depression = 95)



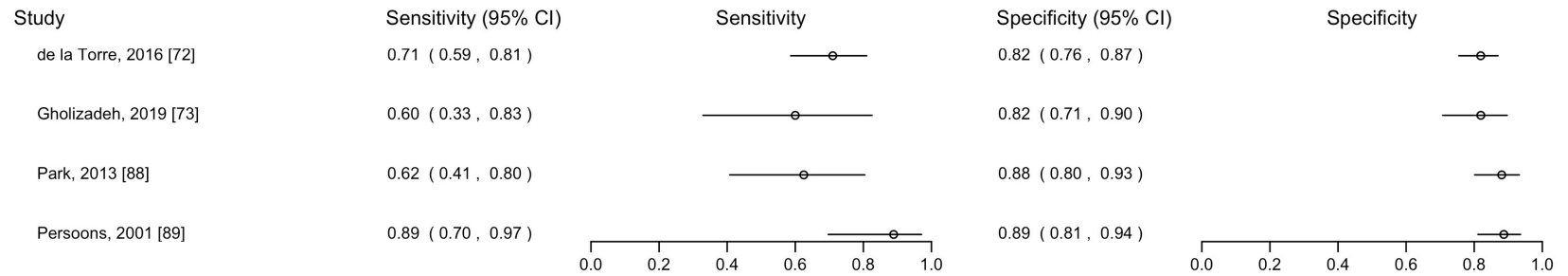
eFigure 2aau. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a non-medical setting, among studies that used the MINI as the reference standard (N Studies = 8; N Participants = 6,792; N major depression = 470)



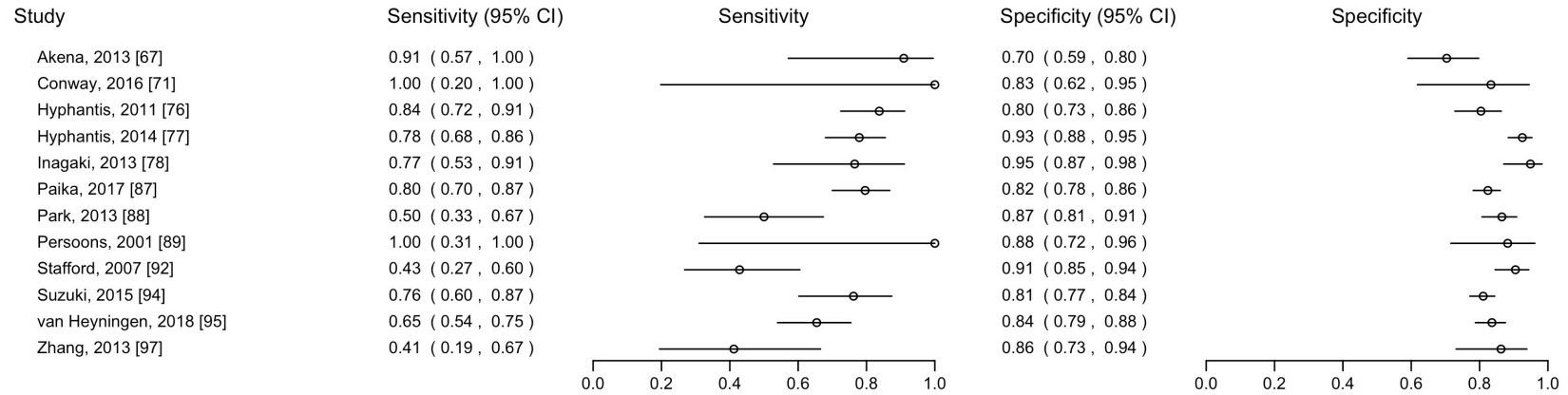
eFigure 2aav. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from a primary care setting, among studies that used the MINI as the reference standard (N Studies = 9; N Participants = 5,092; N major depression = 557)



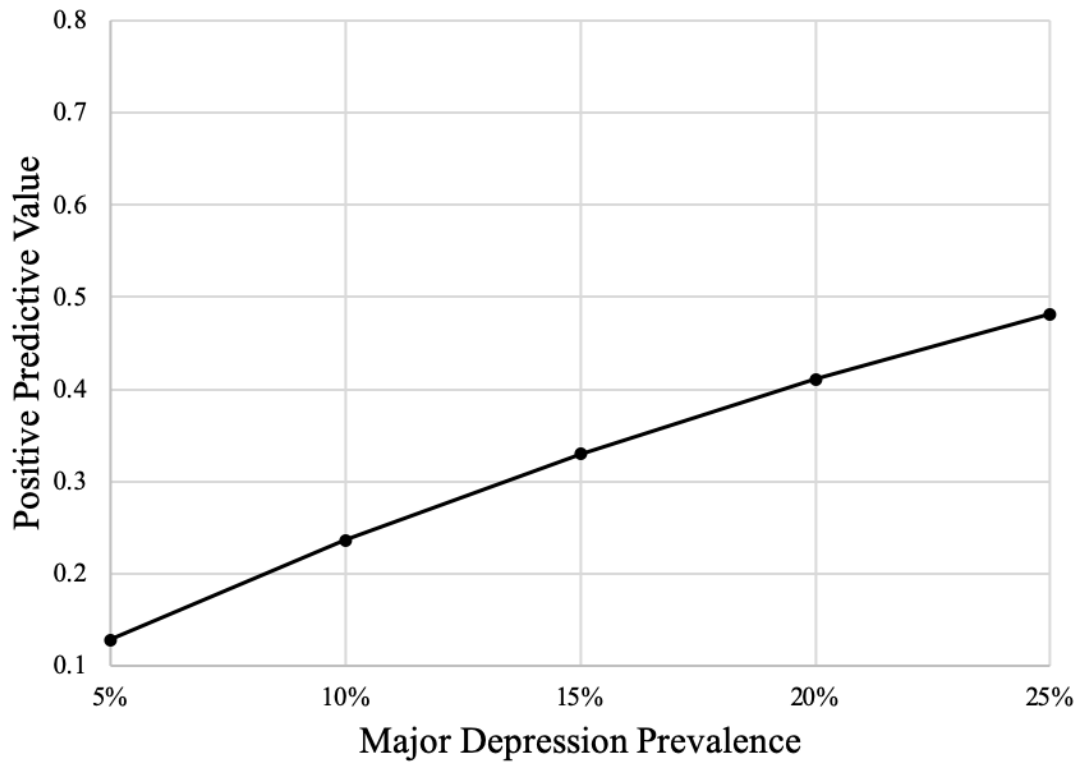
eFigure 2aaw. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from an inpatient specialty care setting, among studies that used the MINI as the reference standard (N Studies = 4; N Participants = 619; N major depression = 135)



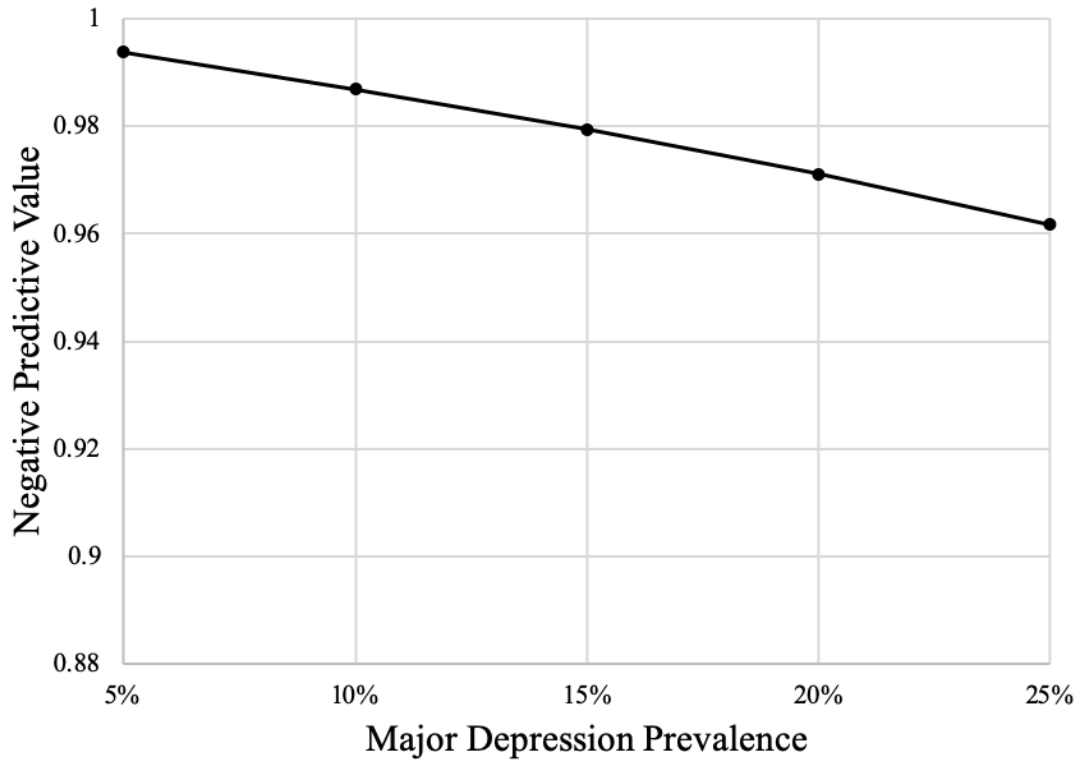
eFigure 2aax. Forest plots of sensitivity and specificity estimates for cutoff 3 of the PHQ-2 among participants from an outpatient specialty care setting, among studies that used the MINI as the reference standard (N Studies = 12; N Participants = 2,663; N major depression = 502)



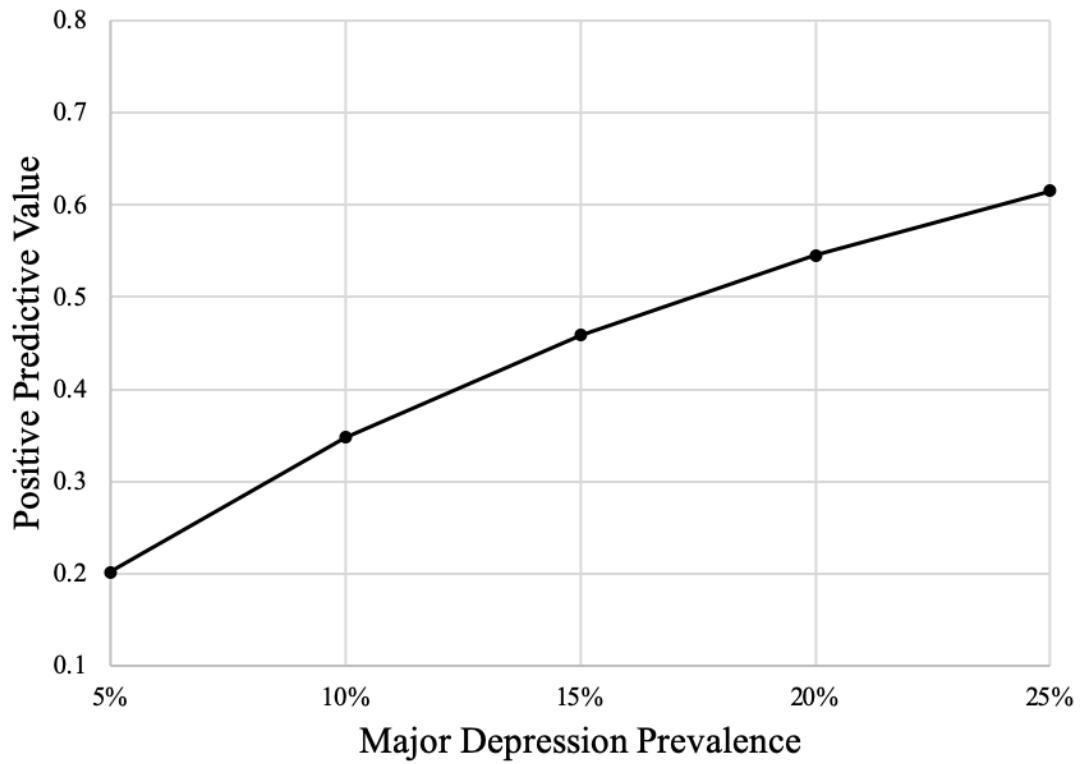
eFigure 3a. Nomogram of positive predictive values for assumed major depression prevalence of 5-25% for PHQ-2 \geq 2, based on accuracy estimates among studies with a semi-structured reference standard and PHQ-9 scores available (N studies = 44, N participants = 10,627, N major depression = 1,361)



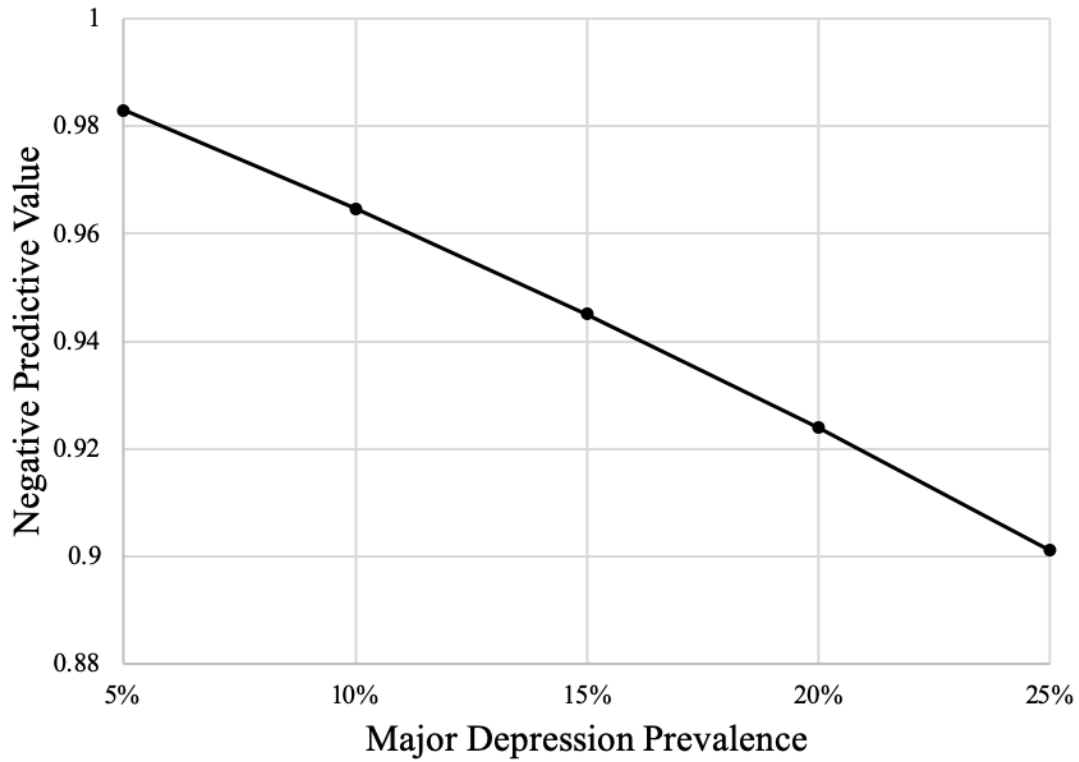
eFigure 3b. Nomogram of negative predictive values for assumed major depression prevalence of 5-25% for PHQ-2 \geq 2, based on accuracy estimates among studies with a semi-structured reference standard and PHQ-9 scores available (N studies = 44, N participants = 10,627, N major depression = 1,361)



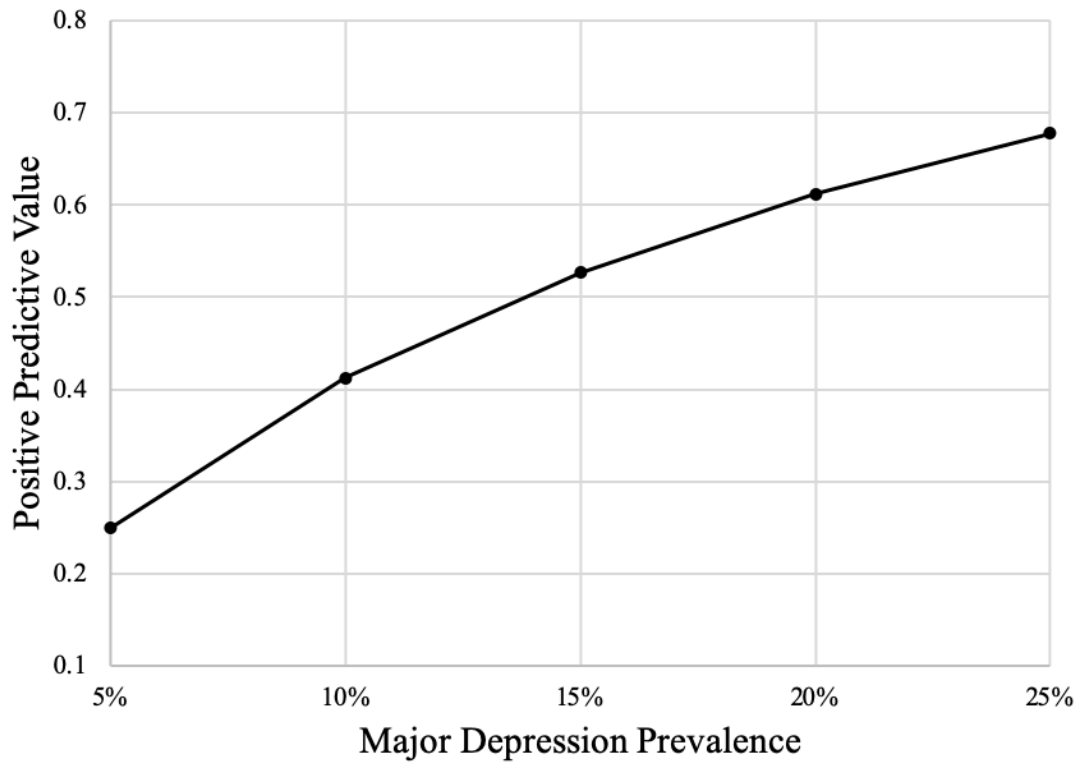
eFigure 3c. Nomogram of positive predictive values for assumed major depression prevalence of 5-25% for PHQ-2 \geq 3, based on accuracy estimates among studies with a semi-structured reference standard and PHQ-9 scores available (N studies = 44, N participants = 10,627, N major depression = 1,361)



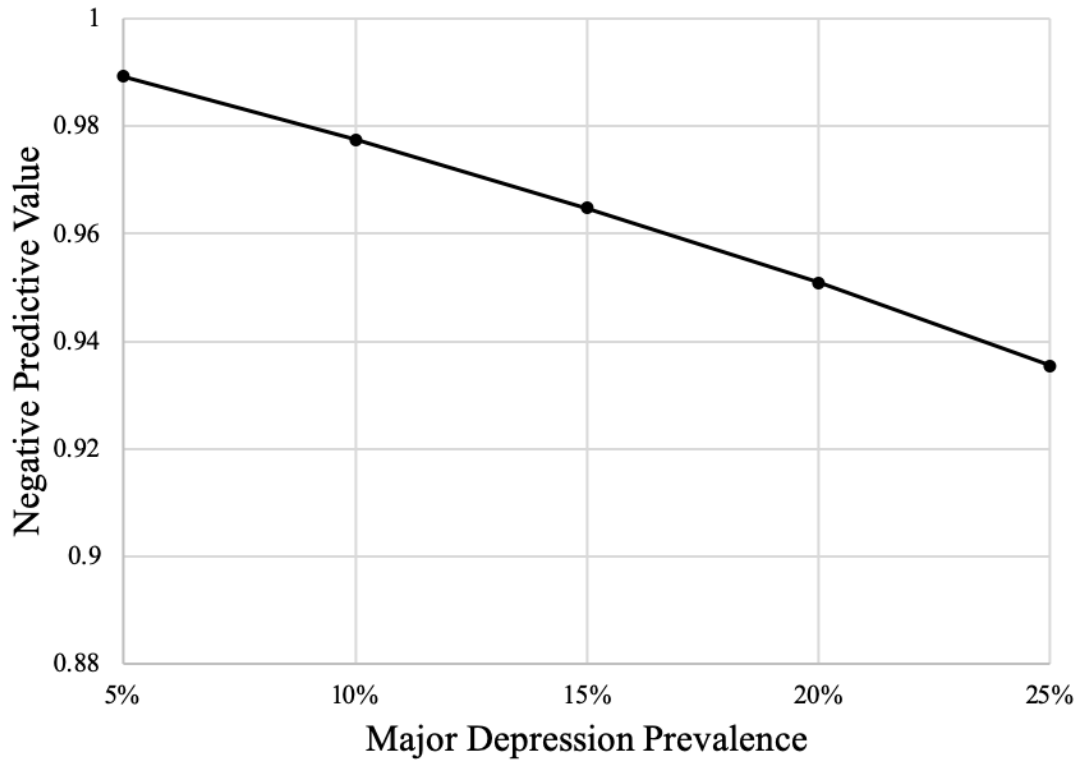
eFigure 3d. Nomogram of negative predictive values for assumed major depression prevalence of 5-25% for PHQ-2 \geq 3, based on accuracy estimates among studies with a semi-structured reference standard and PHQ-9 scores available (N studies = 44, N participants = 10,627, N major depression = 1,361)



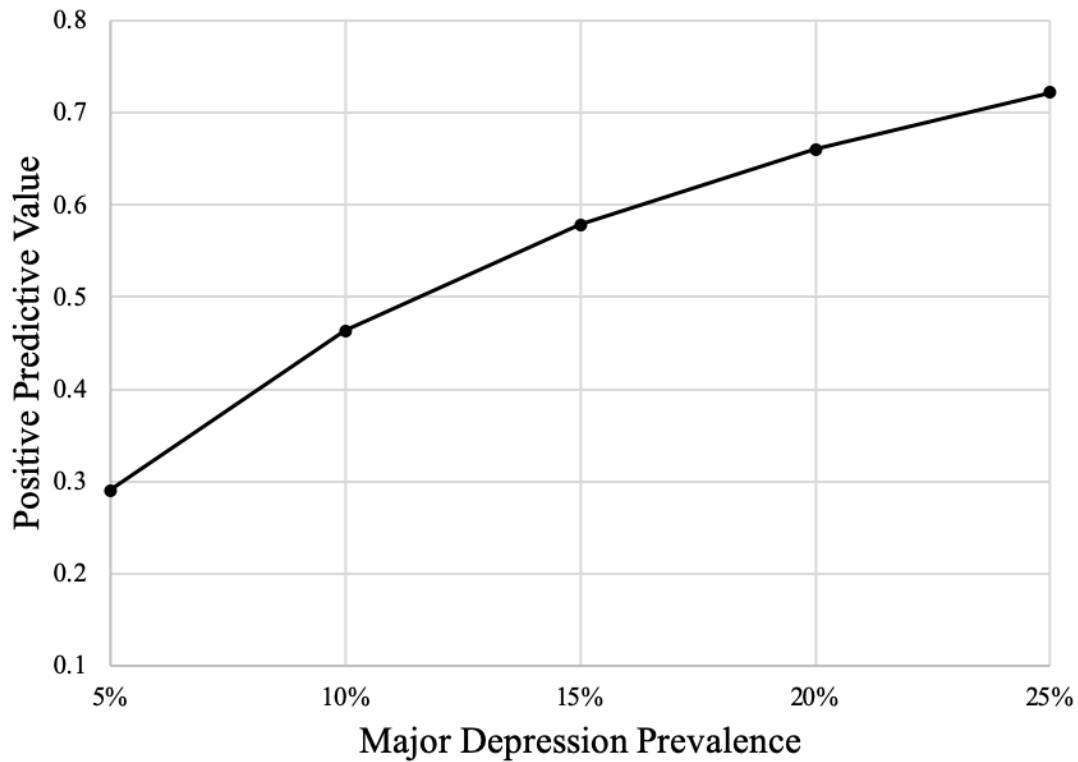
eFigure 3e. Nomogram of positive predictive values for assumed major depression prevalence of 5-25% for PHQ-2 ≥ 2 followed by PHQ-9 ≥ 10 , based on accuracy estimates among studies with a semi-structured reference standard and PHQ-9 scores available (N studies = 44, N participants = 10,627, N major depression = 1,361)



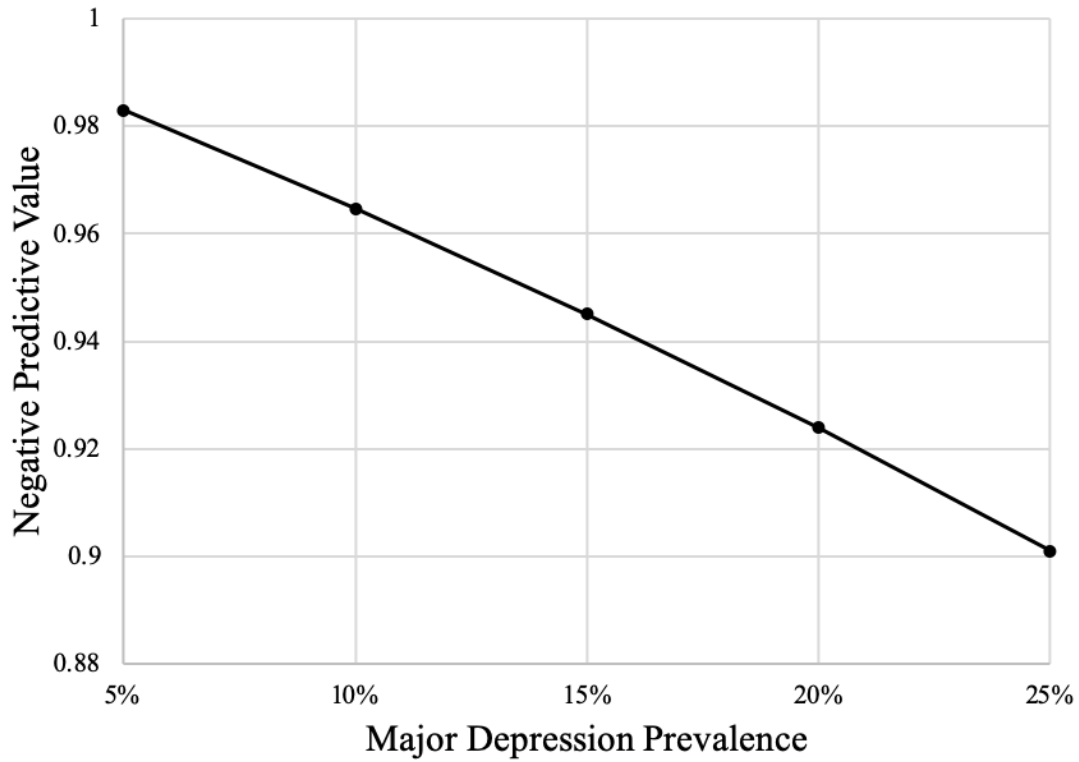
eFigure 3f. Nomogram of negative predictive values for assumed major depression prevalence of 5-25% for PHQ-2 ≥ 2 followed by PHQ-9 ≥ 10 , based on accuracy estimates among studies with a semi-structured reference standard and PHQ-9 scores available (N studies = 44, N participants = 10,627, N major depression = 1,361)



eFigure 3g. Nomogram of positive predictive values for assumed major depression prevalence of 5-25% for PHQ-2 ≥ 3 followed by PHQ-9 ≥ 10 , based on accuracy estimates among studies with a semi-structured reference standard and PHQ-9 scores available (N studies = 44, N participants = 10,627, N major depression = 1,361)



eFigure 3h. Nomogram of negative predictive values for assumed major depression prevalence of 5-25% for PHQ-2 ≥ 3 followed by PHQ-9 ≥ 10 , based on accuracy estimates among studies with a semi-structured reference standard and PHQ-9 scores available (N studies = 44, N participants = 10,627, N major depression = 1,361)



eTable 1a. Characteristics of included primary studies

First Author, Year	Country	Recruited Population	Diagnostic Interview	Classification System	Total N	Major Depression N (%)
Semi-structured Interviews						
Amoozegar, 2017¹	Canada	Migraine patients	SCID	DSM-IV	206	50 (24)
Amtmann, 2015²	USA	Multiple sclerosis patients	SCID	DSM-IV	164	48 (29)
Ayalon, 2010³	Israel	Elderly primary care patients	SCID	DSM-IV	152	6 (4)
Beraldi, 2014⁴	Germany	Cancer inpatients	SCID	DSM-IV	123	9 (7)
Bernstein, 2018⁵	Canada	Inflammatory bowel disease patients	SCID	DSM-IV	242	21 (9)
Bhana, 2015⁶	South Africa	Chronic care patients	SCID	DSM-IV	679	78 (11)
Bombardier, 2012⁷	USA	Inpatients with spinal cord injuries	SCID	DSM-IV	161	14 (9)
Chagas, 2013⁸	Brazil	Outpatients with Parkinson's Disease	SCID	DSM-IV	84	19 (23)
Chibanda, 2016⁹	Zimbabwe	A primary care population with high HIV prevalence	SCID	DSM-IV	264	149 (56)
Eack, 2006¹⁰	USA	Women seeking psychiatric services for their children at two mental health centers	SCID	DSM-IV	49	13 (27)
Fiest, 2014¹¹	Canada	Epilepsy outpatients	SCID	DSM-IV	183	26 (14)
Fischer, 2014¹²	Germany	Heart failure patients	SCID	DSM-IV	194	11 (6)
Gjerdingen, 2009¹³	USA	Mothers registering their newborns for well-child visits at medical or pediatric clinics	SCID	DSM-IV	428	19 (4)
Gräfe, 2004¹⁴	Germany	Medical and psychosomatic outpatients	SCID	DSM-IV	518	71 (14)
Green, 2017¹⁵	USA	Returning veterans	SCID	DSM-V	184	23 (13)
Green, 2018¹⁶	Kenya	Pregnant women and new mothers	SCID	DSM-V	192	10 (5)
Haroz, 2017¹⁷	Myanmar	Primary care patients	SCID	DSM-IV	134	29 (22)
Hitchon, 2019^{18a}	Canada	Rheumatoid arthritis patients	SCID	DSM-IV	152	17 (11)
Khamsch, 2011¹⁹	Iran	Type 2 diabetes patients	SCID	DSM-IV	122	47 (39)
Kwan, 2012²⁰	Singapore	Post-stroke inpatients undergoing rehabilitation	SCID	DSM-IV-TR	114	3 (3)
Lambert, 2015²¹	Australia	Cancer patients	SCID	DSM-IV	161	24 (15)
Lara, 2015²²	Mexico	Pregnant women during the third trimester of pregnancy	SCID	DSM-IV	280	29 (10)
Lino, 2014²³	Brazil	Elderly primary care patients	SCID	DSM-IV	130	31 (24)
Liu, 2011²⁴	Taiwan	Primary care patients	SCAN	DSM-IV	1532	50 (3)
Marrie, 2018²⁵	Canada	Multiple sclerosis patients	SCID	DSM-IV	249	25 (10)
Martin-Subero, 2017²⁶	Spain	Medical inpatients	SCID	DSM-III	1003	83 (8)
McGuire, 2013²⁷	USA	Acute coronary syndrome inpatients	DISH	DSM-IV	100	9 (9)
Osório, 2009²⁸	Brazil	Women in primary care	SCID	DSM-IV	177	60 (34)
Osório, 2012²⁹	Brazil	Inpatients from various clinical wards	SCID	DSM-IV	86	28 (33)

Osório, 2015³⁰	Brazil	Cancer patients	SCID	DSM-IV	399	64 (16)
Patten, 2015³¹	Canada	Multiple sclerosis patients	SCID	DSM-IV	147	20 (14)
Picardi, 2005³²	Italy	Inpatients with skin diseases	SCID	DSM-IV	140	12 (9)
Prisnie, 2016³³	Canada	Stroke and transient ischemic attack patients	SCID	DSM-IV	119	12 (10)
Quinn, Unpublished^a	UK	Stroke patients	SCID	DSM-V	141	17 (12)
Richardson, 2010³⁴	USA	Older adults undergoing in-home aging services care management assessment	SCID	DSM-IV	377	95 (25)
Roch, 2016³⁵	Germany	Inpatient Orthopedic Rehabilitation population	SCID	DSM-IV	42	8 (19)
Rooney, 2013³⁶	UK	Patients with cerebral glioma	SCID	DSM-IV	128	15 (12)
Shinn, 2017³⁷	USA	Cancer patients	SCID	DSM-IV	131	6 (5)
Sidebottom, 2012³⁸	USA	Pregnant women	SCID	DSM-IV	248	12 (5)
Simning, 2012³⁹	USA	Older adults living in public housing	SCID	DSM-IV	190	10 (5)
Spangenberg, 2015⁴⁰	Germany	Primary care patients	SCID	DSM-IV	161	1 (1)
Swartz, 2017⁴¹	Canada	Stroke prevention outpatients	SCID	DSM-IV	379	52 (14)
Turner, 2012⁴²	Australia	Stroke patients	SCID	DSM-IV	72	13 (18)
Turner, Unpublished^a	Australia	Cardiac rehabilitation patients	SCID	DSM-IV	52	4 (8)
Twist, 2013⁴³	UK	Type 2 diabetes outpatients	SCAN	DSM-IV	358	79 (22)
Wagner, 2017⁴⁴	USA	Patients starting radiotherapy for the first diagnosis of any tumor	SCID	DSM-IV	54	6 (11)
Williams, 2012⁴⁵	USA	Parkinson's Disease patients	SCID	DSM-IV	241	65 (27)
Wittkamp, 2009⁴⁶	The Netherlands	Primary care patients at risk for depression	SCID	DSM-IV	261	45 (17)

Fully structured Interviews

Arroll, 2010⁴⁷	New Zealand	Primary care patients	CIDI	DSM-IV	2571	160 (6)
Azah, 2005⁴⁸	Malaysia	Adults attending family medicine clinics	CIDI	ICD-10	180	30 (17)
de Man-van Ginkel, 2012⁴⁹	The Netherlands	Stroke patients	CIDI	DSM-IV	382	54 (14)
Delgadillo, 2011⁵⁰	UK	Injecting drug users	CIS-R	ICD-10	103	51 (50)
Fisher, 2016⁵¹	Australia	Primiparous women less than 6 weeks postpartum	CIDI	DSM-IV	357	4 (1)
Gelaye, 2014⁵²	Ethiopia	Outpatients at a general hospital	CIDI	DSM-IV	923	162 (18)
Grool, 2011⁵³	The Netherlands	Non-demented patients with symptomatic atherosclerotic disease	CIDI	DSM-IV	477	22 (5)
Hahn, 2006⁵⁴	Germany	Patients with chronic illnesses from rehabilitation centers	CIDI	DSM-IV	211	18 (9)
Henkel, 2004⁵⁵	Germany	Primary care patients	CIDI	ICD-10	430	43 (10)
Hobfoll, 2011⁵⁶	Israel	Jewish and Palestinian residents of Jerusalem exposed to war	CIDI	DSM-IV	147	43 (29)
Kiely, 2014⁵⁷	Australia	Community sample of adults	CIDI	ICD-10	823	33 (4)

Kim, 2017 ⁵⁸	South Korea	Randomly selected adults	CIDI	DSM-IV	3076	206 (7)
Kohrt, 2016 ⁵⁹	Nepal	Primary care patients	CIDI	DSM-IV	125	17 (14)
Liu, 2015 ⁶⁰	Canada	Working population	CIDI	DSM-IV	4270	96 (2)
Mohd Sidik, 2012 ⁶¹	Malaysia	Primary care patients	CIDI	DSM-IV	146	31 (21)
Patel, 2008 ⁶²	India	Primary care patients	CIS-R	ICD-10	299	13 (4)
Pence, 2012 ⁶³	Cameroon	HIV-infected patients	CIDI	DSM-IV	398	11 (3)
Razykov, 2013 ⁶⁴	Canada	Patients with systemic sclerosis	CIDI	DSM-IV	345	13 (4)
Thombs, 2008 ⁶⁵	USA	Outpatients with coronary artery disease	C-DIS	DSM-IV	1018	223 (22)
Zuithoff, 2009 ⁶⁶	The Netherlands	General practice patients	CIDI	DSM-IV	1038	135 (13)

Mini International Neuropsychiatric Interviews (MINI)

Akena, 2013 ⁶⁷	Uganda	HIV/AIDS patients	MINI	DSM-IV	92	11 (12)
Baron, 2017 ⁶⁸	South Africa	Xhosa, Afrikaans and Zulu-speaking general population	MINI	DSM-IV	856	93 (11)
Buji, 2018 ⁶⁹	Malaysia	Patients with systemic lupus erythematosus	MINI	DSM-IV	130	5 (4)
Cholera, 2014 ⁷⁰	South Africa	Patients undergoing routine HIV counseling and testing at a primary health care clinic	MINI	DSM-IV	397	47 (12)
Conway, 2016 ⁷¹	Australia	Heart transplant recipients	MINI	DSM-IV	26	2 (8)
de la Torre, 2016 ⁷²	Argentina	Hospitalized general medical patients	MINI	DSM-IV	257	69 (27)
Garabiles, Unpublished ^a	China	Female Filipino domestic workers in Macao	MINI	DSM-IV	99	39 (39)
Gholizadeh, 2019 ^{73a}	Iran	Coronary artery disease patients	MINI	DSM-IV	87	15 (17)
Hantsoo, 2017 ⁷⁴	USA	General population	MINI	DSM-IV	322	19 (6)
Hides, 2007 ⁷⁵	Australia	Injection drug users accessing a needle and syringe program	MINI	DSM-IV	103	47 (46)
Hyphantis, 2011 ⁷⁶	Greece	Patients with various rheumatologic disorders	MINI	DSM-IV	213	69 (32)
Hyphantis, 2014 ⁷⁷	Greece	Patients with chronic illnesses presenting at the emergency department	MINI	DSM-IV	349	95 (27)
Inagaki, 2013 ⁷⁸	Japan	Internal medicine outpatients	MINI	DSM-III-R	104	21 (20)
Janssen, 2016 ⁷⁹	The Netherlands	General population and Type 2 diabetes patients	MINI	DSM-IV	4710	156 (3)
Lamers, 2008 ⁸⁰	The Netherlands	Elderly primary care patients with diabetes mellitus or chronic obstructive pulmonary disease	MINI	DSM-IV	115	65 (57)
Levin-Aspenson, 2017 ⁸¹	USA	General population	MINI	DSM-V	408	66 (16)
Liu, 2016 ⁸²	China	Primary care patients	MINI	DSM-IV	1997	97 (5)
Lotrakul, 2008 ⁸³	Thailand	Outpatients	MINI	DSM-IV	278	19 (7)
Muramatsu, 2007 ⁸⁴	Japan	Primary care patients	MINI	DSM-IV	122	36 (30)
Muramatsu, 2018 ⁸⁵	Japan	Primary care patients	MINI	DSM-IV	163	50 (31)
Nakku, 2016 ⁸⁶	Uganda	Primary patients and hospital outpatients	MINI	DSM-IV	153	84 (55)

Paika, 2017 ⁸⁷	Greece	Patients with long term medical conditions	MINI	DSM-IV	474	98 (21)
Park, 2013 ⁸⁸	South Korea	Cancer patients	MINI	DSM-IV	354	52 (15)
Persoons, 2001 ⁸⁹	Belgium	Inpatients and patients at gastroenterological and hepatology wards	MINI	DSM-IV	179	30 (17)
Rancans, 2018 ⁹⁰	Latvia	Primary care patients	MINI	DSM-IV	1467	147 (10)
Santos, 2013 ⁹¹	Brazil	General population	MINI	DSM-IV	196	25 (13)
Stafford, 2007 ⁹²	Australia	Inpatients with coronary artery disease who had undergone surgery	MINI	DSM-IV	193	35 (18)
Sung, 2013 ⁹³	Singapore	Primary care patients	MINI	DSM-IV	400	12 (3)
Suzuki, 2015 ⁹⁴	Japan	Outpatients in general medicine department	MINI	DSM-IV	512	42 (8)
van Heyningen, 2018 ⁹⁵	South Africa	Pregnant women	MINI	DSM-IV	374	81 (22)
Volker, 2016 ⁹⁶	The Netherlands	Employees on sickness leave	MINI	DSM-IV	98	25 (26)
Zhang, 2013 ⁹⁷	China	Type 2 diabetes patients	MINI	DSM-IV	68	17 (25)

Abbreviations: C-DIS: Computerized Diagnostic Interview Schedule; CIDI: Composite International Diagnostic Interview; CIS-R: Clinical Interview Schedule Revised; DISH: Depression Interview and Structured Hamilton; DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; MINI: Mini Neurosychiatric Diagnostic Interview; SCAN: Schedules for Clinical Assessment in Neuropsychiatry; SCID: Structured Clinical Interview for DSM Disorders; UK: United Kingdom; USA: United States of America.

^aWas unpublished at the time of electronic database search

eTable 1b. Characteristics of eligible primary studies not included in the present study

First Author, Year	Country	Recruited Population	Diagnostic Interview	Classification System	Total N	Major Depression N (%)	Could study have been added as a published dataset? (Reason)
Semi-structured Interviews							
Alamri, 2017 ⁹⁸	Saudi Arabia	Hospitalized elderly in medical and surgical wards	SCID	DSM-IV	199	24 (12)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Bailer, 2016 ⁹⁹	Germany	Healthy participants and cognitive behaviour therapy outpatients	SCID	DSM-IV	200	68 (34)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Becker, 2002 ¹⁰⁰	Saudi Arabia	Primary care patients	SCID	DSM-III-R	173	NR	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Brodey, 2016 ¹⁰¹	USA	Perinatal women	SCID	DSM-IV	879	NR	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Chen, 2013 ¹⁰²	China	Primary care populations	SCID	DSM-IV	280	NR	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Chen, 2012 ¹⁰³	China	Adults over 60 in primary care	SCID	DSM-IV	262	97 (37)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Fann, 2005 ¹⁰⁴	USA	Inpatients with traumatic brain injury	SCID	DSM-IV	135	45 (34)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Fisher, 2016 ¹⁰⁵	USA, Canada	Type 1 diabetes patients	SCID	DSM-V	368	13 (4)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Irmak, 2017 ¹⁰⁶	Turkey	Battered women	SCID	DSM-V	150	63 (42)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Lai, 2010 ¹⁰⁷	China	Men with postpartum wives	SCID	DSM-IV	551	8 (1)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Limon, 2016 ¹⁰⁸	USA	Latino farmworkers	SCID	DSM-IV	99	NR	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Liu, 2016 ¹⁰⁹	China	Rural elderly population	SCID	DSM-IV	839	57 (7)	Yes (Published accuracy results for PHQ-2 cutoffs 1-4)
Nacak, 2017 ¹¹⁰	Germany	Patients with somatoform pain disorder	SCID	DSM-IV	130	36 (28)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Navinés, 2012 ¹¹¹	Spain	Chronic hepatitis C patients	SCID	DSM-IV	500	32 (6)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Phelan, 2010 ¹¹²	USA	Elderly primary care patients	SCID	DSM-IV	69	8 (12)	Yes (Published accuracy results for PHQ-2 cutoffs 1-5)

Thompson, 2011 ¹¹³	USA	Parkinson's patients	SCID	DSM-IV	214	30 (14)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Vöhringer, 2013 ¹¹⁴	Chile	Primary care patients	SCID	DSM-IV	190	59 (31)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Watnick, 2005 ¹¹⁵	USA	Long term dialysis patients	SCID	DSM-IV	62	12 (19)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Fully Structured Interviews							
Al-Ghafri, 2014 ¹¹⁶	Oman	Medical trainees	CIDI	NR	131	NR	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Haddad, 2013 ¹¹⁷	UK	Coronary heart disease patients	CIS-R	ICD-10	730	32 (4)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Ikin, 2016 ¹¹⁸	Australia	Veterans of the Gulf War	CIDI	DSM-IV	1356	NR	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Smith, 2010 ¹¹⁹	USA	Pregnant women	CIDI	DSM-IV	213	13 (6)	No (Published data ineligible; some participants had time intervals between PHQ-2 and CIDI that were greater than 2 weeks).
Valencia-Garcia, 2017 ¹²⁰	USA	Mexican American women	CIDI	DSM-IV	205	40 (20)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Wang, 2015 ¹²¹	China	Cardiovascular outpatients	CIDI	DSM-IV	201	42 (21)	Yes (Published accuracy results for PHQ-2 cutoffs 1-5)
Mini International Neuropsychiatric Interviews (MINI)							
Choi, 2015 ¹²²	Canada	HIV patients	MINI	DSM-IV	190	29 (15)	Yes (Published accuracy results for PHQ-2 cutoffs 2-4)
Du, 2017 ¹²³	China	University students	MINI	DSM-IV	81	9 (11)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Griffith, 2015 ¹²⁴	USA	Patients with epilepsy	MINI	DSM-IV and ICD-10	114	20 (18)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Margrove, 2011 ¹²⁵	UK	Patients with epilepsy	MINI	DSM-IV and ICD-10	52	25 (48)	No (Published accuracy results for PHQ-2 cutoff 3, but results would need to be weighted by inverse selection probabilities)
Persoons, 2003 ¹²⁶	Belgium	Otorhinolaryngology outpatients	MINI	DSM-IV	97	16 (16)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Rathore, 2014 ¹²⁷	USA	Patients with epilepsy	MINI	DSM-IV	158	36 (23)	Yes (Published accuracy results for PHQ-2 cutoffs 1-3)
Scott, 2011 ¹²⁸	USA	Chronic hepatitis C patients	MINI	DSM-IV and ICD-10	30	NR	No (Primary study did not report accuracy results for any PHQ-2 cutoff)

Seo, 2015 ¹²⁹	South Korea	Migrane patients	MINI	DSM-IV	132	39 (30)	Yes (Published accuracy results for PHQ-2 cutoffs 2-4)
van Steenbergen-Weijnenburg, 2010 ¹³⁰	The Netherlands	Diabetes patients	MINI	DSM-IV	197	37 (19)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Wang, 2014 ¹³¹	China	General population	MINI	DSM-IV	1045	28 (3)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Woldetensay, 2018 ¹³²	Ethiopia	Pregnant women	MINI	DSM-IV	216	28 (13)	No (Primary study did not report accuracy results for any PHQ-2 cutoff)
Xiong, 2014 ¹³³	China	Outpatients with multiple somatic symptoms	MINI	DSM-IV	398	116 (29)	Yes (Published accuracy results for PHQ-2 cutoff 3)

Abbreviations: CIDI: Composite International Diagnostic Interview; CIS-R: Clinical Interview Schedule Revised; DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; MINI: Mini International Neuropsychiatric Interview; NR: Not Reported; SCID: Structured Clinical Interview for DSM Disorders; UK: United Kingdom; USA: United States of America.

eTable 2. Numbers of participants and cases of major depression by diagnostic interview

Diagnostic Interview	N Studies	N Participants	Major Depression	
			N	%
Semi-structured				
SCID	45	9,713	1,400	14
SCAN	2	1,890	129	7
DISH	1	100	9	9
Fully structured				
CIDI	17	15,899	1,078	7
DIS	1	1,018	223	22
CIS-R	2	402	64	16
MINI	32	15,296	1,669	11
Total	100	44,318	4,572	10

eTable 3a. Estimates of heterogeneity at PHQ-2 cutoff score of 2

Participant Subgroup	Semi-structured Diagnostic Interviews				Fully Structured Diagnostic Interviews				Mini International Neuropsychiatric Interviews			
	R ^a		τ ²		R ^a		τ ²		R ^a		τ ²	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
All participants	2.72	3.83	1.29	0.27	3.28	9.30	0.64	0.61	3.11	5.37	1.06	0.31
Participants verified to not currently be diagnosed or receiving treatment for a mental health problem	2.05	3.53	1.09	0.43	3.24	9.06	0.96	0.63	2.16	5.31	0.48	0.29
Age < 60	2.62	3.20	1.39	0.23	3.23	8.35	0.76	0.60	2.48	4.52	0.90	0.27
Age ≥ 60	2.20	2.34	1.06	0.20	1.92	4.75	0.43	0.58	2.10	3.24	0.66	0.25
Women	2.68	3.17	1.43	0.26	2.75	7.59	0.65	0.68	2.54	5.53	0.80	0.50
Men	1.66	2.65	0.74	0.26	2.38	5.96	0.60	0.52	2.82	3.94	2.18	0.36
Very high country human development index	1.99	3.47	0.95	0.21	3.42	9.83	0.65	0.63	2.59	5.36	0.76	0.32
High country human development index	3.86	3.60	1.61	0.28	--	--	--	--	2.61	3.26	0.75	0.08
Low-medium country human development index	2.56	3.92	0.37	0.46	2.41	6.47	0.55	0.46	9.11	4.75	4.06	1.26
Non-medical care	1.71	2.86	0.48	0.13	3.38	5.91	0.39	0.11	2.74	6.56	0.63	0.33
Primary care	4.82	5.26	3.00	0.44	2.44	5.38	0.40	0.22	5.01	6.36	2.27	0.31
Inpatient specialty care	2.19	1.88	1.75	0.04	2.59	7.48	0.65	0.92	1.49	1.00	0.30	0.00
Outpatient specialty care	2.01	3.49	0.56	0.28	2.05	9.82	0.25	1.07	1.92	3.30	0.46	0.22

^aR is the ratio of the estimated standard deviation of the pooled sensitivity (or specificity) from the random-effects model to the estimated standard deviation of the pooled sensitivity (or specificity) from the corresponding fixed-effects model

eTable 3b. Estimates of heterogeneity at PHQ-2 cutoff score of 3

Participant Subgroup	Semi-structured Diagnostic Interviews				Fully Structured Diagnostic Interviews				Mini International Neuropsychiatric Interviews			
	R ^a		τ ²		R ^a		τ ²		R ^a		τ ²	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
All participants	2.22	3.47	0.47	0.33	3.50	7.00	0.59	0.78	2.94	4.13	0.56	0.40
Participants verified to not currently be diagnosed or receiving treatment for a mental health problem	2.37	3.11	0.73	0.58	2.48	5.12	0.32	0.48	2.37	4.4	0.41	0.54
Age < 60	2.10	3.02	0.48	0.29	3.08	6.75	0.55	0.88	2.55	3.60	0.55	0.37
Age ≥ 60	2.23	2.20	0.95	0.24	2.66	3.81	0.92	0.81	2.32	2.83	0.52	0.49
Women	2.34	3.18	0.74	0.39	2.79	5.87	0.56	0.91	2.66	4.49	0.58	0.74
Men	1.35	2.07	0.18	0.2	2.46	4.63	0.52	0.7	2.13	2.87	0.71	0.41
Very high country human development index	1.74	3.09	0.28	0.24	3.54	6.88	0.59	0.77	2.68	4.07	0.44	0.44
High country human development index	3.68	2.56	1.19	0.19	--	--	--	--	1.97	3.22	0.21	0.15
Low-medium country human development index	2.56	3.92	0.37	0.46	3.98	4.12	0.93	0.25	5.60	3.34	2.42	0.81
Non-medical care	1.56	2.25	0.05	0.11	2.01	6.49	0.13	0.44	2.84	5.83	0.45	0.87
Primary care	3.22	4.39	0.82	0.44	2.73	4.89	0.34	0.38	3.78	4.26	0.84	0.28
Inpatient specialty care	1.76	1.12	0.48	0.00	1.67	4.81	0.15	0.54	1.22	1.00	0.05	0.00
Outpatient specialty care	1.74	3.30	0.27	0.41	3.08	7.39	0.46	1.07	2.40	2.45	0.44	0.18

^aR is the ratio of the estimated standard deviation of the pooled sensitivity (or specificity) from the random-effects model to the estimated standard deviation of the pooled sensitivity (or specificity) from the corresponding fixed-effects model

eTable 4a. Comparison of PHQ-2 sensitivity and specificity estimates at cutoff 2 among all participants, among participants not currently diagnosed or receiving treatment for a mental health problem, and among participant subgroups based on age, sex, human development index, and care setting

Participant Subgroup	Semi-structured Diagnostic Interviews				Fully Structured Diagnostic Interviews				Mini International Neuropsychiatric Interviews			
	Sensitivity		Specificity		Sensitivity		Specificity		Sensitivity		Specificity	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
All participants	0.91	(0.88, 0.94)	0.67	(0.64, 0.71)	0.82	(0.75, 0.87)	0.71	(0.63, 0.77)	0.89	(0.84, 0.92)	0.68	(0.64, 0.73)
Participants verified to not currently be diagnosed or receiving treatment for a mental health problem	0.90	(0.84, 0.94)	0.73	(0.67, 0.78)	0.91	(0.78, 0.97)	0.71	(0.54, 0.83)	0.85	(0.78, 0.90)	0.74	(0.68, 0.79)
Age < 60	0.93	(0.89, 0.96)	0.64	(0.61, 0.68)	0.82	(0.74, 0.87)	0.70	(0.62, 0.77)	0.89	(0.85, 0.93)	0.67	(0.62, 0.71)
Age ≥ 60	0.89	(0.81, 0.93)	0.73	(0.69, 0.76)	0.84	(0.74, 0.90)	0.74	(0.65, 0.81)	0.87	(0.80, 0.92)	0.73	(0.68, 0.77)
Women	0.93	(0.88, 0.95)	0.65	(0.61, 0.68)	0.80	(0.72, 0.86)	0.68	(0.52, 0.72)	0.88	(0.84, 0.92)	0.64	(0.58, 0.70)
Men	0.89	(0.84, 0.93)	0.71	(0.67, 0.75)	0.82	(0.74, 0.88)	0.73	(0.65, 0.79)	0.92	(0.85, 0.96)	0.72	(0.67, 0.76)
Very high country human development index	0.93	(0.89, 0.95)	0.68	(0.65, 0.72)	0.82	(0.74, 0.88)	0.72	(0.63, 0.79)	0.91	(0.86, 0.94)	0.69	(0.64, 0.74)
High country human development index	0.86	(0.70, 0.94)	0.68	(0.59, 0.76)	--	--	--	--	0.84	(0.73, 0.91)	0.65	(0.60, 0.70)
Low-medium country human development index	0.87	(0.69, 0.95)	0.50	(0.31, 0.69)	0.76	(0.57, 0.88)	0.65	(0.49, 0.79)	0.87	(0.15, 1.00)	0.73	(0.35, 0.93)
Non-medical care	0.90	(0.71, 0.97)	0.66	(0.53, 0.77)	0.69	(0.53, 0.82)	0.79	(0.73, 0.85)	0.85	(0.75, 0.92)	0.71	(0.62, 0.79)
Primary care	0.95	(0.86, 0.98)	0.69	(0.61, 0.76)	0.87	(0.78, 0.93)	0.66	(0.58, 0.74)	0.94	(0.84, 0.98)	0.62	(0.52, 0.70)
Inpatient specialty care	0.93	(0.81, 0.98)	0.62	(0.58, 0.66)	0.88	(0.61, 0.97)	0.62	(0.29, 0.86)	0.89	(0.78, 0.95)	0.69	(0.65, 0.73)
Outpatient specialty care	0.89	(0.83, 0.93)	0.68	(0.61, 0.73)	0.81	(0.72, 0.87)	0.72	(0.54, 0.85)	0.87	(0.80, 0.91)	0.70	(0.64, 0.76)

Abbreviations: CI: confidence interval

eTable 4b. Comparison of PHQ-2 sensitivity and specificity estimates at cutoff 3 among all participants, among participants not currently diagnosed or receiving treatment for a mental health problem, and among participant subgroups based on age, sex, human development index, and care setting

Participant Subgroup	Semi-structured Diagnostic Interviews				Fully Structured Diagnostic Interviews				Mini International Neuropsychiatric Interviews			
	Sensitivity		Specificity		Sensitivity		Specificity		Sensitivity		Specificity	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
All participants	0.72	(0.67, 0.77)	0.85	(0.83, 0.87)	0.53	(0.44, 0.62)	0.89	(0.84, 0.92)	0.69	(0.62, 0.75)	0.87	(0.84, 0.90)
Participants verified to not currently be diagnosed or receiving treatment for a mental health problem	0.68	(0.57, 0.77)	0.90	(0.86, 0.92)	0.55	(0.40, 0.68)	0.91	(0.83, 0.95)	0.67	(0.58, 0.75)	0.89	(0.85, 0.93)
Age < 60	0.71	(0.66, 0.77)	0.84	(0.81, 0.86)	0.52	(0.43, 0.61)	0.89	(0.84, 0.93)	0.69	(0.62, 0.75)	0.87	(0.84, 0.89)
Age ≥ 60	0.77	(0.67, 0.84)	0.88	(0.86, 0.90)	0.59	(0.43, 0.73)	0.88	(0.82, 0.93)	0.71	(0.62, 0.79)	0.87	(0.83, 0.90)
Women	0.74	(0.67, 0.80)	0.84	(0.81, 0.87)	0.51	(0.42, 0.61)	0.88	(0.78, 0.88)	0.68	(0.60, 0.74)	0.85	(0.81, 0.89)
Men	0.70	(0.64, 0.75)	0.87	(0.85, 0.89)	0.55	(0.45, 0.65)	0.89	(0.85, 0.93)	0.73	(0.64, 0.80)	0.88	(0.85, 0.91)
Very high country human development index	0.73	(0.68, 0.77)	0.87	(0.84, 0.89)	0.52	(0.42, 0.62)	0.91	(0.86, 0.94)	0.73	(0.66, 0.79)	0.87	(0.84, 0.90)
High country human development index	0.73	(0.55, 0.86)	0.83	(0.77, 0.87)	--	--	--	--	0.60	(0.51, 0.69)	0.87	(0.83, 0.90)
Low-medium country human development index	0.63	(0.44, 0.79)	0.69	(0.50, 0.83)	0.61	(0.34, 0.82)	0.81	(0.72, 0.88)	0.62	(0.13, 0.95)	0.86	(0.60, 0.96)
Non-medical care	0.77	(0.62, 0.88)	0.83	(0.74, 0.89)	0.37	(0.27, 0.47)	0.94	(0.89, 0.97)	0.63	(0.50, 0.74)	0.90	(0.83, 0.95)
Primary care	0.75	(0.64, 0.84)	0.86	(0.81, 0.90)	0.67	(0.54, 0.77)	0.87	(0.80, 0.91)	0.67	(0.52, 0.80)	0.86	(0.80, 0.90)
Inpatient specialty care	0.75	(0.61, 0.84)	0.81	(0.79, 0.83)	0.76	(0.58, 0.88)	0.80	(0.58, 0.92)	0.72	(0.62, 0.80)	0.85	(0.81, 0.88)
Outpatient specialty care	0.69	(0.62, 0.75)	0.87	(0.83, 0.90)	0.43	(0.30, 0.56)	0.89	(0.79, 0.95)	0.72	(0.62, 0.80)	0.86	(0.82, 0.89)

Abbreviations: CI: confidence interval

eTable 4c. Comparison of PHQ-2 sensitivity and specificity estimates among participants verified to not currently be diagnosed or receiving treatment for a mental health problem compared to all participants, among participants administered a semi-structured diagnostic interview

Cutoff	All participants ^a				Participants verified to not currently be diagnosed or receiving treatment for a mental health problem ^b				Difference across groups (All participants – participants verified to not currently be diagnosed or receiving treatment for a mental health problem)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.91	(0.88, 0.94)	0.67	(0.64, 0.71)	0.90	(0.84, 0.94)	0.73	(0.67, 0.78)	0.01	(-0.05, 0.08)	-0.06	(-0.10, -0.02)
3	0.72	(0.67, 0.77)	0.85	(0.83, 0.87)	0.68	(0.57, 0.77)	0.90	(0.86, 0.92)	0.04	(-0.05, 0.16)	-0.04	(-0.07, -0.02)

^aN Studies = 48; N Participants = 11,703; N major depression = 1,538; AUC (95% CI) = 0.875 (0.864, 0.887)

^bN Studies = 25; N Participants = 3,708; N major depression = 527; AUC (95% CI) = 0.895 (0.876, 0.913)

Abbreviations: CI: confidence interval

eTable 4d. Comparison of PHQ-2 sensitivity and specificity estimates among participants aged < 60 compared to ≥ 60, among participants administered a semi-structured diagnostic interview

Cutoff	Age < 60 ^a				Age ≥ 60 ^b				Difference across groups (Age < 60 – Age ≥ 60)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.93	(0.89, 0.96)	0.64	(0.61, 0.68)	0.89	(0.81, 0.93)	0.73	(0.69, 0.76)	0.04	(-0.05, 0.08)	-0.08	(-0.13, -0.05)
3	0.71	(0.66, 0.77)	0.84	(0.81, 0.86)	0.77	(0.67, 0.84)	0.88	(0.86, 0.90)	-0.05	(-0.22, 0.03)	-0.04	(-0.07, -0.01)

^aN Studies = 43; N Participants = 7,759; N major depression = 1117; AUC (95% CI) = 0.867 (0.853, 0.881)

^bN Studies = 40; N Participants = 3,875; N major depression = 415; AUC (95% CI) = 0.886 (0.865, 0.908)

Abbreviations: CI: confidence interval

eTable 4e. Comparison of PHQ-2 sensitivity and specificity estimates among women compared to men, among participants administered a semi-structured diagnostic interview

Cutoff	Women ^a				Men ^b				Difference across groups (Women – Men)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.93	(0.88, 0.95)	0.65	(0.61, 0.68)	0.89	(0.84, 0.93)	0.71	(0.67, 0.75)	0.03	(-0.04, 0.08)	-0.06	(-0.10, -0.03)
3	0.74	(0.67, 0.80)	0.84	(0.81, 0.87)	0.70	(0.64, 0.75)	0.87	(0.85, 0.89)	0.04	(-0.06, 0.12)	-0.03	(-0.05, 0.00)

^aN Studies = 47; N Participants = 7,280; N major depression = 1,054; AUC (95% CI) = 0.871 (0.857, 0.885)

^bN Studies = 40; N Participants = 4,345; N major depression = 484; AUC (95% CI) = 0.880 (0.859, 0.900)

Abbreviations: CI: confidence interval

Table 4f. Comparison of PHQ-2 sensitivity and specificity estimates among participants from countries with a very high human development index compared to a high human development index, among participants administered a semi-structured diagnostic interview

Cutoff	Very high human development index ^a				High human development index ^b				Difference across groups ^c (Very high human development index – high human development index)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.93	(0.89, 0.95)	0.68	(0.65, 0.72)	0.86	(0.70, 0.94)	0.68	(0.59, 0.76)	0.06	(-0.05, 0.23)	0.00	(-0.08, 0.10)
3	0.73	(0.68, 0.77)	0.87	(0.84, 0.89)	0.73	(0.55, 0.86)	0.83	(0.77, 0.87)	0.01	(-0.17, 0.18)	0.04	(-0.01, 0.09)

^aN Studies = 37; N Participants = 9,156; N major depression = 944; AUC (95% CI) = 0.890 (0.877, 0.904)

^bN Studies = 8; N Participants = 1,957; N major depression = 356; AUC (95% CI) = 0.861 (0.835, 0.886)

^c2 bootstrap iterations (0.2%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4g. Comparison of PHQ-2 sensitivity and specificity estimates among participants from countries with a very high human development index compared to a low-medium human development index, among participants administered a semi-structured diagnostic interview

Cutoff	Very high human development index ^a				Low-medium human development index ^b				Difference across groups ^c (Very high human development index – low-medium human development index)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.93	(0.89, 0.95)	0.68	(0.65, 0.72)	0.87	(0.69, 0.95)	0.50	(0.31, 0.69)	0.06	(-0.03, 0.17)	0.18	(0.06, 0.33)
3	0.73	(0.68, 0.77)	0.87	(0.84, 0.89)	0.63	(0.44, 0.79)	0.69	(0.50, 0.83)	0.10	(-0.08, 0.27)	0.18	(0.08, 0.32)

^aN Studies = 37; N Participants = 9,156; N major depression = 944; AUC (95% CI) = 0.890 (0.877, 0.904)

^bN Studies = 3; N Participants = 590; N major depression = 188; AUC (95% CI) = 0.737 (0.691, 0.782)

^c309 bootstrap iterations (30.9%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4h. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and non-medical care settings, among participants administered a semi-structured diagnostic interview

Cutoff	Primary care ^a				Non-medical care ^b				Difference across groups ^c (Primary care – non-medical care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.95	(0.86, 0.98)	0.69	(0.61, 0.76)	0.90	(0.71, 0.97)	0.66	(0.53, 0.77)	0.05	(-0.04, 0.14)	0.03	(-0.04, 0.12)
3	0.75	(0.64, 0.84)	0.86	(0.81, 0.90)	0.77	(0.62, 0.88)	0.83	(0.74, 0.89)	-0.02	(-0.16, 0.15)	0.03	(-0.02, 0.08)

^aN Studies = 15; N Participants = 4,569; N major depression = 667; AUC (95% CI) = 0.890 (0.874, 0.907)

^bN Studies = 2; N Participants = 567; N major depression = 105; AUC (95% CI) = 0.882 (0.839, 0.925)

^c577 bootstrap iterations (57.7%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4i. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and inpatient specialty care settings, among participants administered a semi-structured diagnostic interview

Cutoff	Primary care ^a				Inpatient specialty care ^b				Difference across groups ^c (Primary care – inpatient specialty care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.95	(0.86, 0.98)	0.69	(0.61, 0.76)	0.93	(0.81, 0.98)	0.62	(0.58, 0.66)	0.02	(-0.08, 0.15)	0.07	(-0.01, 0.16)
3	0.75	(0.64, 0.84)	0.86	(0.81, 0.90)	0.75	(0.61, 0.84)	0.81	(0.79, 0.83)	0.01	(-0.17, 0.22)	0.05	(0.00, 0.11)

^aN Studies = 15; N Participants = 4,569; N major depression = 667; AUC (95% CI) = 0.890 (0.874, 0.907)

^bN Studies = 10; N Participants = 2,019; N major depression = 184; AUC (95% CI) = 0.862 (0.828, 0.897)

^c1 bootstrap iteration (0.1%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4j. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and outpatient speciality care settings, among participants administered a semi-structured diagnostic interview

Cutoff	Primary care ^a				Outpatient specialty care ^b				Difference across groups (Primary care – outpatient specialty care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.95	(0.86, 0.98)	0.69	(0.61, 0.76)	0.89	(0.83, 0.93)	0.68	(0.62, 0.73)	0.06	(-0.03, 0.12)	0.01	(-0.07, 0.09)
3	0.75	(0.64, 0.84)	0.86	(0.81, 0.90)	0.69	(0.62, 0.75)	0.87	(0.83, 0.90)	0.06	(-0.07, 0.18)	-0.01	(-0.06, 0.04)

^aN Studies = 15; N Participants = 4,569; N major depression = 667; AUC (95% CI) = 0.890 (0.874, 0.907)

^bN Studies = 23; N Participants = 4,548; N major depression = 582; AUC (95% CI) = 0.875 (0.856, 0.893)

Abbreviations: CI: confidence interval

eTable 4k. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 1 (Participant Selection) - Signalling Question 1 (*Was a consecutive or random sample of participants enrolled?*), among participants administered a semi-structured diagnostic interview

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups ^c (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.94	(0.81, 0.99)	0.66	(0.54, 0.76)	0.91	(0.87, 0.94)	0.67	(0.64, 0.71)	0.04	(-0.10, 0.09)	-0.01	(-0.15, 0.09)
3	0.75	(0.62, 0.85)	0.87	(0.79, 0.92)	0.71	(0.66, 0.77)	0.85	(0.82, 0.87)	0.04	(-0.13, 0.18)	0.02	(-0.07, 0.07)

^aN Studies = 7; N Participants = 1,416; N major depression = 203

^bN Studies = 41; N Participants = 10,287; N major depression = 1,335

^c7 bootstrap iterations (0.7%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4l. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 3 (Reference Standard) - Signalling Question 2 (*Were the reference standard results interpreted without knowledge of the results of the index test?*) , among participants administered a semi-structured diagnostic interview

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.92	(0.88, 0.95)	0.63	(0.59, 0.67)	0.90	(0.81, 0.96)	0.73	(0.68, 0.77)	0.02	(-0.06, 0.11)	-0.09	(-0.15, -0.04)
3	0.73	(0.67, 0.79)	0.83	(0.79, 0.86)	0.70	(0.60, 0.78)	0.88	(0.85, 0.90)	0.04	(-0.08, 0.15)	-0.05	(-0.09, -0.01)

^aN Studies = 28; N Participants = 7,068; N major depression = 932

^bN Studies = 20; N Participants = 4,635; N major depression = 606

Abbreviations: CI: confidence interval

eTable 4m. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 3 (Reference Standard) - Signalling Question 3 (*Did a qualified person administer the reference standard?*), among participants administered a semi-structured diagnostic interview

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.90	(0.85, 0.93)	0.68	(0.63, 0.72)	0.94	(0.87, 0.98)	0.66	(0.60, 0.72)	-0.04	(-0.11, 0.04)	0.01	(-0.05, 0.08)
3	0.70	(0.64, 0.75)	0.85	(0.82, 0.88)	0.77	(0.66, 0.85)	0.85	(0.82, 0.88)	-0.07	(-0.18, 0.06)	0.00	(-0.04, 0.05)

^aN Studies = 34; N Participants = 8,648; N major depression = 1,052

^bN Studies = 14; N Participants = 3,055; N major depression = 486

Abbreviations: CI: confidence interval

eTable 4n. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “unclear” risk of bias for QUADAS-2 Domain 4 (Flow and Timing) - Signalling Question 1 (*Was there an appropriate interval between index test and reference standard?*), among participants administered a semi-structured interview

Cutoff	Low risk of bias ^a				Unclear risk of bias ^b				Difference across groups (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.92	(0.88, 0.95)	0.66	(0.62, 0.70)	0.90	(0.78, 0.96)	0.71	(0.66, 0.76)	0.02	(-0.06, 0.12)	-0.06	(-0.12, 0.00)
3	0.74	(0.69, 0.79)	0.84	(0.81, 0.86)	0.64	(0.52, 0.74)	0.89	(0.86, 0.92)	0.11	(-0.05, 0.24)	-0.05	(-0.09, 0.01)

^aN Studies = 40; N Participants = 9,382; N major depression = 1260

^bN Studies = 13; N Participants = 2,294; N major depression = 278

Abbreviations: CI: confidence interval

eTable 4o. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 4 (Flow and Timing) - Signalling Question 4 (*Were all patients included in the analysis?*), among participants administered a semi-structured diagnostic interview

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.88	(0.83, 0.92)	0.67	(0.62, 0.71)	0.96	(0.91, 0.98)	0.68	(0.62, 0.73)	-0.07	(-0.13, 0.00)	-0.01	(-0.07, 0.05)
3	0.71	(0.64, 0.77)	0.84	(0.81, 0.87)	0.75	(0.67, 0.81)	0.87	(0.83, 0.90)	-0.04	(-0.13, 0.10)	-0.02	(-0.06, 0.02)

^aN Studies = 33; N Participants = 6,874; N major depression = 1,057

^bN Studies = 15; N Participants = 4,829; N major depression = 481

Abbreviations: CI: confidence interval

eTable 4p. Comparison of PHQ-2 sensitivity and specificity estimates among participants verified to not currently be diagnosed or receiving treatment for a mental health problem compared to all participants, among participants administered a fully structured diagnostic interview

Cutoff	All participants ^a				Participants verified to not currently be diagnosed or receiving treatment for a mental health problem ^b				Difference across groups ^c (All participants – participants verified to not currently be diagnosed or receiving treatment for a mental health problem)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.82	(0.75, 0.87)	0.71	(0.63, 0.77)	0.91	(0.78, 0.97)	0.71	(0.54, 0.83)	-0.09	(-0.19, 0.00)	0.00	(-0.10, 0.16)
3	0.53	(0.44, 0.62)	0.89	(0.84, 0.92)	0.55	(0.40, 0.68)	0.91	(0.83, 0.95)	-0.02	(-0.17, 0.14)	-0.02	(-0.06, 0.07)

^aN Studies = 20; N Participants = 17,319; N major depression = 1,365; AUC (95% CI) = 0.821 (0.807, 0.835)

^bN Studies = 5; N Participants = 4,050; N major depression = 292; AUC (95% CI) = 0.875 (0.848, 0.901)

^c54 bootstrap iterations (5.4%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4q. Comparison of PHQ-2 sensitivity and specificity estimates among participants aged < 60 compared to ≥ 60, among participants administered a fully structured diagnostic interview

Cutoff	Age < 60 ^a				Age ≥ 60 ^b				Difference across groups (Age < 60 – Age ≥ 60)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.82	(0.74, 0.87)	0.70	(0.62, 0.77)	0.84	(0.74, 0.90)	0.74	(0.65, 0.81)	-0.02	(-0.13, 0.10)	-0.04	(-0.10, 0.04)
3	0.52	(0.43, 0.61)	0.89	(0.84, 0.93)	0.59	(0.43, 0.73)	0.88	(0.82, 0.93)	-0.06	(-0.24, 0.12)	0.01	(-0.02, 0.05)

^aN Studies = 20; N Participants = 13,901; N major depression = 1,097; AUC (95% CI) = 0.816 (0.800, 0.832)

^bN Studies = 15; N Participants = 3,400; N major depression = 268; AUC (95% CI) = 0.841 (0.811, 0.871)

Abbreviations: CI: confidence interval

eTable 4r. Comparison of PHQ-2 sensitivity and specificity estimates among women compared to men, among participants administered a fully structured diagnostic interview

Cutoff	Women ^a				Men ^b				Difference across groups (Women – Men)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.80	(0.72, 0.86)	0.68	(0.52, 0.72)	0.82	(0.74, 0.88)	0.73	(0.65, 0.79)	-0.02	(-0.13, 0.07)	-0.05	(-0.10, -0.01)
3	0.51	(0.42, 0.61)	0.88	(0.78, 0.88)	0.55	(0.45, 0.65)	0.89	(0.85, 0.93)	-0.04	(-0.17, 0.05)	-0.01	(-0.04, 0.02)

^aN Studies = 20; N Participants = 9,690; N major depression = 802; AUC (95% CI) = 0.809 (0.791, 0.828)

^bN Studies = 18; N Participants = 7,619; N major depression = 561; AUC (95% CI) = 0.827 (0.806, 0.849)

Abbreviations: CI: confidence interval

eTable 4s. Comparison of PHQ-2 sensitivity and specificity estimates among participants from countries with a very high human development index compared to a low-medium human development index, among participants administered a fully structured diagnostic interview

Cutoff	Very high human development index ^a				Low-medium human development index ^b				Difference across groups ^c (Very high human development index – low-medium human development index)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.82	(0.74, 0.88)	0.72	(0.63, 0.79)	0.76	(0.57, 0.88)	0.65	(0.49, 0.79)	0.07	(-0.14, 0.22)	0.06	(-0.08, 0.21)
3	0.52	(0.42, 0.62)	0.91	(0.86, 0.94)	0.61	(0.34, 0.82)	0.81	(0.72, 0.88)	-0.09	(-0.42, 0.16)	0.10	(0.02, 0.17)

^aN Studies = 16; N Participants = 15,574; N major depression = 1,162; AUC (95% CI) = 0.833 (0.819, 0.848)

^bN Studies = 4; N Participants = 1,745; N major depression = 203; AUC (95% CI) = 0.751 (0.711, 0.791)

^c131 bootstrap iterations (13.1%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4t. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and non-medical care settings, among participants administered a fully structured diagnostic interview

Cutoff	Primary care ^a				Non-medical care ^b				Difference across groups ^c (Primary care – non-medical care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.87	(0.78, 0.93)	0.66	(0.58, 0.74)	0.69	(0.53, 0.82)	0.79	(0.73, 0.85)	0.18	(-0.01, 0.36)	-0.13	(-0.22, -0.02)
3	0.67	(0.54, 0.77)	0.87	(0.80, 0.91)	0.37	(0.27, 0.47)	0.94	(0.89, 0.97)	0.30	(0.11, 0.46)	-0.07	(-0.14, -0.01)

^aN Studies = 7; N Participants = 4,789; N major depression = 429; AUC (95% CI) = 0.844 (0.820, 0.867)

^bN Studies = 4; N Participants = 8,316; N major depression = 378; AUC (95% CI) = 0.801 (0.774, 0.829)

^c119 bootstrap iterations (11.9%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4u. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and inpatient specialty care settings, among participants administered a fully structured diagnostic interview

Cutoff	Primary care ^a				Inpatient specialty care ^b				Difference across groups ^c (Primary care – inpatient specialty care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.87	(0.78, 0.93)	0.66	(0.58, 0.74)	0.88	(0.61, 0.97)	0.62	(0.29, 0.86)	-0.01	(-0.15, 0.13)	0.05	(-0.05, 0.15)
3	0.67	(0.54, 0.77)	0.87	(0.80, 0.91)	0.76	(0.58, 0.88)	0.80	(0.58, 0.92)	-0.10	(-0.30, 0.05)	0.07	(-0.01, 0.12)

^aN Studies = 7; N Participants = 4,789; N major depression = 429; AUC (95% CI) = 0.844 (0.820, 0.867)

^bN Studies = 2; N Participants = 593; N major depression = 72; AUC (95% CI) = 0.824 (0.764, 0.885)

^c582 bootstrap iterations (58.2%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4v. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and outpatient specialty care settings, among participants administered a fully structured diagnostic interview

Cutoff	Primary care ^a				Outpatient specialty care ^b				Difference across groups ^c (Primary care – outpatient specialty care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.87	(0.78, 0.93)	0.66	(0.58, 0.74)	0.81	(0.72, 0.87)	0.72	(0.54, 0.85)	0.06	(-0.06, 0.23)	-0.05	(-0.19, 0.13)
3	0.67	(0.54, 0.77)	0.87	(0.80, 0.91)	0.43	(0.30, 0.56)	0.89	(0.79, 0.95)	0.24	(-0.02, 0.46)	-0.02	(-0.11, 0.08)

^aN Studies = 7; N Participants = 4,789; N major depression = 429; AUC (95% CI) = 0.844 (0.820, 0.867)

^bN Studies = 7; N Participants = 3,621; N major depression = 486; AUC (95% CI) = 0.808 (0.784, 0.832)

^c6 bootstrap iterations (0.6%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

Table 4w. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 1 (Participant Selection) - Signalling Question 1 (*Was a consecutive or random sample of participants enrolled?*), among participants administered a fully structured diagnostic interview

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups ^c (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.91	(0.75, 0.97)	0.72	(0.60, 0.81)	0.80	(0.72, 0.86)	0.70	(0.60, 0.78)	0.11	(-0.11, 0.20)	0.02	(-0.12, 0.13)
3	0.67	(0.49, 0.81)	0.87	(0.80, 0.91)	0.50	(0.40, 0.60)	0.90	(0.84, 0.93)	0.17	(-0.11, 0.38)	-0.03	(-0.12, 0.03)

^aN Studies = 5; N Participants = 3,539; N major depression = 232

^bN Studies = 15; N Participants = 13,780; N major depression = 1,133

^c49 bootstrap iterations (4.9%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4x. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 3 (Reference Standard) - Signalling Question 2 (*Were the reference standard results interpreted without knowledge of the results of the index test?*), among participants administered a fully structured diagnostic interview

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.81	(0.70, 0.88)	0.74	(0.67, 0.80)	0.82	(0.71, 0.89)	0.64	(0.48, 0.78)	-0.01	(-0.16, 0.11)	0.10	(-0.04, 0.28)
3	0.56	(0.44, 0.67)	0.91	(0.86, 0.94)	0.51	(0.36, 0.65)	0.86	(0.76, 0.93)	0.05	(-0.15, 0.25)	0.05	(-0.03, 0.15)

^aN Studies = 12; N Participants = 10,020; N major depression = 927

^bN Studies = 8; N Participants = 7,299; N major depression = 438

Abbreviations: CI: confidence interval

eTable 4y. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “unclear” risk of bias for QUADAS-2 Domain 4 (Flow and Timing) - Signalling Question 1 (*Was there an appropriate interval between index test and reference standard?*), among participants administered a fully structured interview

Cutoff	Low risk of bias ^a				Unclear risk of bias ^b				Difference across groups ^c (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.82	(0.74, 0.88)	0.70	(0.61, 0.77)	0.79	(0.65, 0.88)	0.75	(0.61, 0.85)	0.03	(-0.13, 0.17)	-0.06	(-0.18, 0.09)
3	0.56	(0.46, 0.65)	0.88	(0.83, 0.92)	0.45	(0.27, 0.65)	0.91	(0.83, 0.95)	0.10	(-0.12, 0.33)	-0.03	(-0.09, 0.05)

^aN Studies = 15; N Participants = 11,311; N major depression = 1,145

^bN Studies = 6; N Participants = 6,008; N major depression = 220

^c8 bootstrap iterations (0.8%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4z. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 4 (Flow and Timing) - Signalling Question 4 (*Were all patients included in the analysis?*), among participants administered a fully structured diagnostic interview

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups ^c (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.82	(0.74, 0.88)	0.70	(0.61, 0.78)	0.78	(0.65, 0.87)	0.71	(0.61, 0.79)	0.04	(-0.10, 0.17)	-0.01	(-0.10, 0.09)
3	0.54	(0.43, 0.65)	0.88	(0.83, 0.92)	0.48	(0.38, 0.58)	0.93	(0.90, 0.95)	0.06	(-0.11, 0.21)	-0.05	(-0.09, 0.00)

^aN Studies = 17; N Participants = 15,278; N major depression = 1,167

^bN Studies = 3; N Participants = 2,041; N major depression = 198

^c305 bootstrap iterations (30.5%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4aa. Comparison of PHQ-2 sensitivity and specificity estimates among participants verified to not currently be diagnosed or receiving treatment for a mental health problem compared to all participants, among participants administered the MINI

Cutoff	All participants ^a				Participants verified to not currently be diagnosed or receiving treatment for a mental health problem ^b				Difference across groups (All participants – participants verified to not currently be diagnosed or receiving treatment for a mental health problem)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.89	(0.84, 0.92)	0.68	(0.64, 0.73)	0.85	(0.78, 0.90)	0.74	(0.68, 0.79)	0.04	(-0.02, 0.12)	-0.05	(-0.10, 0.00)
3	0.69	(0.62, 0.75)	0.87	(0.84, 0.90)	0.67	(0.58, 0.75)	0.89	(0.85, 0.93)	0.01	(-0.07, 0.12)	-0.02	(-0.05, 0.01)

^aN Studies = 32; N Participants = 15,296; N major depression = 1,669; AUC (95% CI) = 0.866 (0.854, 0.877)

^bN Studies = 15; N Participants = 8390; N major depression = 581; AUC (95% CI) = 0.863 (0.844, 0.882)

Abbreviations: CI: confidence interval

eTable 4ab. Comparison of PHQ-2 sensitivity and specificity estimates among participants aged < 60 compared to ≥ 60, among participants administered the MINI

Cutoff	Age < 60 ^a				Age ≥ 60 ^b				Difference across groups (Age < 60 – Age ≥ 60)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.89	(0.85, 0.93)	0.67	(0.62, 0.71)	0.87	(0.80, 0.92)	0.73	(0.68, 0.77)	0.02	(-0.06, 0.09)	-0.06	(-0.12, 0.00)
3	0.69	(0.62, 0.75)	0.87	(0.84, 0.89)	0.71	(0.62, 0.79)	0.87	(0.83, 0.90)	-0.02	(-0.17, 0.07)	0.00	(-0.04, 0.03)

^aN Studies = 31; N Participants = 10,071; N major depression = 1,153; AUC (95% CI) = 0.863 (0.849, 0.877)

^bN Studies = 26; N Participants = 5,192; N major depression = 506; AUC (95% CI) = 0.878 (0.858, 0.897)

Abbreviations: CI: confidence interval

eTable 4ac. Comparison of PHQ-2 sensitivity and specificity estimates among women compared to men, among participants administered the MINI

Cutoff	Women ^a				Men ^b				Difference across groups (Women – Men)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.88	(0.84, 0.92)	0.64	(0.58, 0.70)	0.92	(0.85, 0.96)	0.72	(0.67, 0.76)	-0.04	(-0.09, 0.01)	-0.08	(-0.14, 0.03)
3	0.68	(0.60, 0.74)	0.85	(0.81, 0.89)	0.73	(0.64, 0.80)	0.88	(0.85, 0.91)	-0.05	(-0.15, 0.05)	-0.02	(-0.06, 0.01)

^aN Studies = 31; N Participants = 9,053; N major depression = 1,138; AUC (95% CI) = 0.849 (0.834, 0.863)

^bN Studies = 29; N Participants = 6,225; N major depression = 530; AUC (95% CI) = 0.895 (0.877, 0.913)

Abbreviations: CI: confidence interval

eTable 4ad. Comparison of PHQ-2 sensitivity and specificity estimates among participants from countries with a very high human development index compared to a high human development index, among participants administered the MINI

Cutoff	Very high human development index ^a				High human development index ^b				Difference across groups ^c (Very high human development index – high human development index)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.91	(0.86, 0.94)	0.69	(0.64, 0.74)	0.84	(0.73, 0.91)	0.65	(0.60, 0.70)	0.07	(-0.02, 0.19)	0.04	(0.04, 0.13)
3	0.73	(0.66, 0.79)	0.87	(0.84, 0.90)	0.60	(0.51, 0.69)	0.87	(0.83, 0.90)	0.13	(0.02, 0.25)	0.01	(-0.04, 0.05)

^aN Studies = 21; N Participants = 10,699; N major depression = 1,141; AUC (95% CI) = 0.885 (0.872, 0.898)

^bN Studies = 9; N Participants = 4,352; N major depression = 433; AUC (95% CI) = 0.826 (0.801,0.850)

Abbreviations: CI: confidence interval

eTable 4ae. Comparison of PHQ-2 sensitivity and specificity estimates among participants from countries with a very high human development index compared to a low-medium human development index, among participants administered the MINI

Cutoff	Very high human development index ^a				Low-medium human development index ^b				Difference across groups ^c (Very high human development index – low-medium human development index)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.91	(0.86, 0.94)	0.69	(0.64, 0.74)	0.87	(0.15, 1.00)	0.73	(0.35, 0.93)	0.04	(-0.04, 0.11)	-0.04	(-0.12, 0.04)
3	0.73	(0.66, 0.79)	0.87	(0.84, 0.90)	0.62	(0.13, 0.95)	0.86	(0.60, 0.96)	0.11	(-0.14, 0.32)	0.02	(-0.04, 0.08)

^aN Studies = 21; N Participants = 10,699; N major depression = 1,141; AUC (95% CI) = 0.885 (0.872, 0.898)

^bN Studies = 2; N Participants = 245; N major depression = 95; AUC (95% CI) = 0.824 (0.767, 0.881)

^c591 bootstrap iterations (59.1%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4af. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and non-medical care settings, among participants administered the MINI

Cutoff	Primary care ^a				Non-medical care ^b				Difference across groups ^c (Primary care – non-medical care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.94	(0.84, 0.98)	0.62	(0.52, 0.70)	0.85	(0.75, 0.92)	0.71	(0.62, 0.79)	0.09	(-0.03, 0.23)	-0.10	(-0.22, 0.03)
3	0.67	(0.52, 0.80)	0.86	(0.80, 0.90)	0.63	(0.50, 0.74)	0.90	(0.83, 0.95)	0.05	(-0.14, 0.28)	-0.05	(-0.12, 0.02)

^aN Studies = 9; N Participants = 5,092; N major depression = 557; AUC (95% CI) = 0.861 (0.841, 0.881)

^bN Studies = 8; N Participants = 6,792; N major depression = 470; AUC (95% CI) = 0.862 (0.841, 0.884)

^c3 bootstrap iterations (0.3%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4ag. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and inpatient specialty care settings, among participants administered the MINI

Cutoff	Primary care ^a				Inpatient specialty care ^b				Difference across groups ^c (Primary care – inpatient specialty care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.94	(0.84, 0.98)	0.62	(0.52, 0.70)	0.89	(0.78, 0.95)	0.69	(0.65, 0.73)	0.05	(-0.07, 0.18)	-0.07	(-0.20, 0.04)
3	0.67	(0.52, 0.80)	0.86	(0.80, 0.90)	0.72	(0.62, 0.80)	0.85	(0.81, 0.88)	-0.05	(-0.24, 0.17)	0.01	(-0.07, 0.06)

^aN Studies = 9; N Participants = 5,092; N major depression = 557; AUC (95% CI) = 0.861 (0.841, 0.881)

^bN Studies = 4; N Participants = 619; N major depression = 135; AUC (95% CI) = 0.868 (0.828, 0.908)

^c123 bootstrap iterations (12.3%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4ah. Comparison of PHQ-2 sensitivity and specificity estimates among participants from primary care and outpatient specialty care settings, among participants administered the MINI

Cutoff	Primary care ^a				Outpatient specialty care ^b				Difference across groups ^c (Primary care – outpatient specialty care)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.94	(0.84, 0.98)	0.62	(0.52, 0.70)	0.87	(0.80, 0.91)	0.70	(0.64, 0.76)	0.07	(-0.04, 0.16)	-0.09	(-0.21, 0.03)
3	0.67	(0.52, 0.80)	0.86	(0.80, 0.90)	0.72	(0.62, 0.80)	0.86	(0.82, 0.89)	-0.05	(-0.23, 0.14)	0.00	(-0.07, 0.05)

^aN Studies = 9; N Participants = 5,092; N major depression = 557; AUC (95% CI) = 0.861 (0.841, 0.881)

^bN Studies = 12; N Participants = 2,663; N major depression = 502; AUC (95% CI) = 0.867 (0.847, 0.888)

^c2 bootstrap iterations (0.2%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4ai. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 1 (Participant Selection) - Signalling Question 1 (*Was a consecutive or random sample of participants enrolled?*), among participants administered the MINI

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.89	(0.84, 0.93)	0.68	(0.62, 0.74)	0.89	(0.80, 0.94)	0.68	(0.62, 0.74)	0.01	(-0.07, 0.10)	0.00	(-0.08, 0.08)
3	0.71	(0.64, 0.78)	0.89	(0.85, 0.92)	0.66	(0.55, 0.75)	0.85	(0.81, 0.89)	0.05	(-0.08, 0.20)	0.03	(-0.01, 0.08)

^aN Studies = 14; N Participants = 11,244; N major depression = 897

^bN Studies = 18; N Participants = 4,052; N major depression = 772

Abbreviations: CI: confidence interval

eTable 4aj. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 3 (Reference Standard) - Signalling Question 2 (*Were the reference standard results interpreted without knowledge of the results of the index test?*), among participants administered the MINI

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.89	(0.83, 0.94)	0.67	(0.61, 0.72)	0.89	(0.81, 0.93)	0.71	(0.64, 0.77)	0.01	(-0.06, 0.11)	-0.04	(-0.11, 0.04)
3	0.70	(0.61, 0.78)	0.86	(0.83, 0.89)	0.67	(0.59, 0.74)	0.89	(0.83, 0.93)	0.03	(-0.06, 0.19)	-0.03	(-0.07, 0.03)

^aN Studies = 22; N Participants = 6,252; N major depression = 1,067

^bN Studies = 10; N Participants = 9,044; N major depression = 602

Abbreviations: CI: confidence interval

eTable 4ak. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “unclear” risk of bias for QUADAS-2 Domain 4 (Flow and Timing) - Signalling Question 1 (*Was there an appropriate interval between index test and reference standard?*), among participants administered the MINI

Cutoff	Low risk of bias ^a				Unclear risk of bias ^b				Difference across groups (Low risk of bias – unclear risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.89	(0.83, 0.93)	0.69	(0.64, 0.74)	0.90	(0.79, 0.95)	0.67	(0.56, 0.76)	-0.01	(-0.10, 0.12)	0.02	(-0.08, 0.17)
3	0.70	(0.62, 0.77)	0.87	(0.84, 0.90)	0.64	(0.52, 0.75)	0.87	(0.77, 0.93)	0.05	(-0.16, 0.25)	0.00	(-0.06, 0.10)

^aN Studies = 28; N Participants = 12,090; N major depression = 1,410

^bN Studies = 8; N Participants = 3,187; N major depression = 259

Abbreviations: CI: confidence interval

eTable 4a. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “unclear” risk of bias for QUADAS-2 Domain 4 (Flow and Timing) - Signalling Question 2 (Did all patients receive a reference standard?), among participants administered the MINI

Cutoff	Low risk of bias ^a				Unclear risk of bias ^b				Difference across groups ^c (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.90	(0.85, 0.94)	0.68	(0.64, 0.72)	0.81	(0.63, 0.92)	0.72	(0.53, 0.86)	0.09	(-0.03, 0.26)	-0.04	(-0.17, 0.14)
3	0.69	(0.62, 0.75)	0.87	(0.84, 0.90)	0.64	(0.43, 0.81)	0.87	(0.78, 0.92)	0.05	(-0.16, 0.26)	0.00	(-0.06, 0.08)

^aN Studies = 27; N Participants = 14,151; N major depression = 1,429

^bN Studies = 5; N Participants = 1,145; N major depression = 240

^c48 bootstrap iterations (4.8%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 4am. Comparison of PHQ-2 sensitivity and specificity estimates among studies and participants categorized as having “low” risk of bias compared to “high” or “unclear” risk of bias for QUADAS-2 Domain 4 (Flow and Timing) - Signalling Question 4 (*Were all patients included in the analysis?*), among participants administered the MINI

Cutoff	Low risk of bias ^a				Unclear or high risk of bias ^b				Difference across groups ^c (Low risk of bias – unclear or high risk of bias)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
2	0.89	(0.85, 0.93)	0.67	(0.62, 0.71)	0.86	(0.62, 0.95)	0.77	(0.63, 0.87)	0.04	(-0.09, 0.26)	-0.11	(-0.21, 0.11)
3	0.71	(0.65, 0.77)	0.86	(0.83, 0.88)	0.52	(0.34, 0.71)	0.92	(0.82, 0.97)	0.19	(-0.10, 0.40)	-0.06	(-0.11, 0.05)

^aN Studies = 27; N Participants = 9,912; N major depression = 1260

^bN Studies = 5; N Participants = 5,384; N major depression = 409

^c60 bootstrap iterations (6.0%) did not produce a difference estimate at cutoffs 2 and 3. These iterations were removed prior to determining the bootstrapped CIs for the respective cutoffs.

Abbreviations: CI: confidence interval

eTable 5. Sensitivity and specificity estimates for the PHQ-2 alone, the PHQ-9 alone, and for PHQ-2 \geq 2 followed by PHQ-9, among 44 studies (N participants = 10,627; N major depression = 1,361) that used a semi-structured reference standard and had both PHQ-2 and PHQ-9 item scores available

PHQ-2 alone			PHQ-9 alone			PHQ-2 \geq 2 followed by PHQ-9		
cutoff	Sensitivity	Specificity	cutoff	Sensitivity	Specificity	cutoff	Sensitivity	Specificity
0	1.00	0.00	0	1.00	0.00	0	1.00	0.00
1	0.98	0.46	1	1.00	0.12	1	0.92	0.67
2	0.92	0.67	2	1.00	0.22	2	0.92	0.67
3	0.72	0.85	3	0.99	0.33	3	0.92	0.67
4	0.56	0.93	4	0.99	0.44	4	0.92	0.68
5	0.35	0.97	5	0.98	0.53	5	0.92	0.70
6	0.22	0.99	6	0.97	0.61	6	0.91	0.73
			7	0.96	0.68	7	0.90	0.76
			8	0.93	0.74	8	0.88	0.79
			9	0.89	0.79	9	0.85	0.83
			10	0.86	0.85	10	0.82	0.87
			11	0.82	0.88	11	0.79	0.89
			12	0.75	0.90	12	0.73	0.91
			13	0.67	0.93	13	0.66	0.93
			14	0.62	0.94	14	0.60	0.95
			15	0.53	0.96	15	0.52	0.96
			16	0.45	0.97	16	0.44	0.97
			17	0.37	0.98	17	0.36	0.98
			18	0.31	0.98	18	0.31	0.98
			19	0.27	0.99	19	0.26	0.99
			20	0.22	0.99	20	0.21	0.99
			21	0.17	1.00	21	0.17	1.00
			22	0.13	1.00	22	0.13	1.00
			23	0.10	1.00	23	0.10	1.00
			24	0.07	1.00	24	0.07	1.00
			25	0.05	1.00	25	0.05	1.00
			26	0.03	1.00	26	0.03	1.00
			27	0.02	1.00	27	0.02	1.00

eTable 6. Comparison of sensitivity and specificity for PHQ-2 \geq 2 in combination with PHQ-9 \geq 5 to 15 versus sensitivity and specificity for PHQ-9 \geq 5 to 15, among studies that used a semi-structured diagnostic interview as the reference standard (N Studies = 44; N Participants = 10,627; N major depression = 1,361)

Cutoff	PHQ-2 \geq 2 followed by PHQ-9 \geq 5 to 15				PHQ-9 \geq 5 to 15			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
5	0.92	(0.88, 0.95)	0.70	(0.67, 0.73)	0.98	(0.95, 0.99)	0.53	(0.48, 0.57)
6	0.91	(0.87, 0.94)	0.73	(0.69, 0.76)	0.97	(0.94, 0.98)	0.61	(0.57, 0.65)
7	0.90	(0.86, 0.94)	0.76	(0.72, 0.79)	0.96	(0.92, 0.98)	0.68	(0.64, 0.72)
8	0.88	(0.83, 0.92)	0.79	(0.76, 0.82)	0.93	(0.88, 0.95)	0.74	(0.70, 0.77)
9 ^a	0.85	(0.80, 0.89)	0.83	(0.80, 0.85)	0.89	(0.85, 0.93)	0.79	(0.76, 0.82)
10	0.82	(0.76, 0.86)	0.87	(0.84, 0.89)	0.86	(0.80, 0.90)	0.85	(0.82, 0.87)
11	0.79	(0.73, 0.84)	0.89	(0.87, 0.91)	0.82	(0.76, 0.87)	0.88	(0.85, 0.90)
12	0.73	(0.67, 0.78)	0.91	(0.89, 0.93)	0.75	(0.69, 0.81)	0.90	(0.88, 0.92)
13	0.66	(0.60, 0.71)	0.93	(0.91, 0.94)	0.67	(0.61, 0.73)	0.93	(0.91, 0.94)
14	0.60	(0.54, 0.66)	0.95	(0.93, 0.96)	0.62	(0.55, 0.68)	0.94	(0.93, 0.96)
15	0.52	(0.45, 0.58)	0.96	(0.94, 0.97)	0.53	(0.46, 0.59)	0.96	(0.94, 0.97)

^aFor PHQ-2 \geq 2 in combination with PHQ-9 \geq 9, the default optimizer in glmer failed, thus bobyqa was used instead

Abbreviations: CI: confidence interval

eTable 7. QUADAS-2 ratings for each primary study included in the present study

First Author, Year	Domain 1: Participant Selection					Domain 2: Index Test				Domain 3: Reference Standard					Domain 4: Flow and Timing				
	SQ 1	SQ2	SQ3	RoB	AC	SQ 1	SQ2	RoB	AC	SQ 1	SQ2	SQ3	RoB	AC	SQ 1	SQ2	SQ3	SQ4	RoB
Semi-structured Interviews																			
Amoozegar, 2017 ¹	U/C	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	U/C	U/C	Low	U/C	Yes	Yes	No	U/C
Amtmann, 2015 ²	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Ayalon, 2010 ³	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Beraldi, 2014 ⁴	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	U/C	U/C	Low	Yes	Yes	Yes	Yes	Low
Bernstein, 2018 ⁵	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Bhana, 2015 ⁶	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	U/C
Bombardier, 2012 ⁷	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	IPD ^b	Yes	Yes	U/C	U/C
Chagas, 2013 ⁸	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	No	U/C
Chibanda, 2016 ⁹	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Eack, 2006 ¹⁰	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Fiest, 2014 ¹¹	U/C	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	U/C	U/C	Low	U/C	Yes	Yes	No	U/C
Fischer, 2014 ¹²	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Gjerdingen, 2009 ¹³	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	U/C	U/C
Gräfe, 2004 ¹⁴	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	U/C	U/C	Low	Yes	Yes	Yes	U/C	U/C
Green, 2017 ¹⁵	No	Yes	No	High	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	IPD ^b	Yes	Yes	No	U/C
Green, 2018 ¹⁶	U/C	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Haroz, 2017 ¹⁷	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Hitchon, 2019 ^{18a}	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	U/C	Yes	Yes	Yes	U/C
Khamseh, 2011 ¹⁹	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Kwan, 2012 ²⁰	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	U/C	U/C	U/C	Yes	Yes	Yes	U/C	U/C
Lambert, 2015 ²¹	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Lara, 2015 ²²	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Lino, 2014 ²³	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	U/C	Yes	Yes	Yes	U/C
Liu, 2011 ²⁴	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	No	U/C
Marrie, 2018 ²⁵	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Martin-Subero, 2017 ²⁶	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
McGuire, 2013 ²⁷	U/C	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	No	High	Low	Yes	Yes	Yes	Yes	Low
Osório, 2009 ²⁸	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	U/C	U/C	Low	Yes	Yes	Yes	Yes	Low
Osório, 2012 ²⁹	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	U/C	U/C	Low	Yes	Yes	Yes	Yes	Low
Osório, 2015 ³⁰	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	U/C	U/C	Low	Yes	Yes	Yes	Yes	Low
Patten, 2015 ³¹	U/C	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	U/C	Yes	Yes	Yes	U/C
Picardi, 2005 ³²	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Prisnie, 2016 ³³	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	U/C	Yes	Yes	No	U/C
Quinn, Unpublished ^a	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Richardson, 2010 ³⁴	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	U/C	U/C	Low	Yes	Yes	Yes	Yes	Low
Roch, 2016 ³⁵	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	U/C	U/C	Low	Yes	Yes	Yes	Yes	Low
Rooney, 2013 ³⁶	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Shinn, 2017 ³⁷	U/C	Yes	Yes	Low	High	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Sidebottom, 2012 ³⁸	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	IPD ^b	Yes	Yes	No	U/C

Simning, 2012³⁹	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	U/C	U/C	Low	Yes	Yes	Yes	Yes	Low
Spangenberg, 2015⁴⁰	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	U/C	Yes	U/C	No	High	U/C	Yes	Yes	Yes	Yes	Low
Swartz, 2017⁴¹	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	No	No	High	Low	IPD ^b	Yes	Yes	No	High
Turner, 2012⁴²	U/C	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Turner, Unpublished^a	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Twist, 2013⁴³	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	No	Yes	High	Low	Yes	Yes	Yes	U/C	U/C
Wagner, 2017⁴⁴	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	IPD ^b	No	Yes	No	High
Williams, 2012⁴⁵	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	IPD ^b	Yes	Yes	Yes	IPD ^b
Wittkampf, 2009⁴⁶	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	No	U/C
Fully Structured Interviews																			
Arroll, 2010⁴⁷	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Azah, 2005⁴⁸	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	U/C	Yes	U/C	Yes	U/C	U/C
de Man-van Ginkel, 2012⁴⁹	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Delgado, 2011⁵⁰	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Fisher, 2016⁵¹	U/C	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Gelaye, 2014⁵²	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Grool, 2011⁵³	U/C	Yes	Yes	Low	Low	N/A	N/A	Low	U/C	Yes	U/C	Yes	U/C	Low	U/C	Yes	Yes	Yes	U/C
Hahn, 2006⁵⁴	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	U/C	Yes	Yes	Yes	U/C
Henkel, 2004⁵⁵	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Hobfoll, 2011⁵⁶	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	U/C	Yes	Yes	Yes	U/C
Kiely, 2014⁵⁷	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	U/C	U/C	Yes	U/C	U/C
Kim, 2017⁵⁸	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Kohrt, 2016⁵⁹	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Liu, 2015⁶⁰	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	U/C	Yes	Yes	Yes	U/C
Mohd Sidik, 2012⁶¹	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	U/C	Yes	Yes	Yes	Yes	Low
Patel, 2008⁶²	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Pence, 2012⁶³	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Razykov, 2013⁶⁴	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Thombs, 2008⁶⁵	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Zuithoff, 2009⁶⁶	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	U/C	Yes	Yes	Yes	Low	U/C	IPD ^b	Yes	Yes	No	U/C
Mini International Neuropsychiatric Interviews (MINI)																			
Akena, 2013⁶⁷	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Baron, 2017⁶⁸	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Buji, 2018⁶⁹	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Cholera, 2014⁷⁰	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	U/C	Yes	Yes	Yes	Low	U/C	Yes	No	Yes	Yes	High
Conway, 2016⁷¹	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	IPD ^b	Yes	Yes	Yes	IPD ^b
de la Torre, 2016⁷²	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Garabiles, Unpublished^a	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Gholizadeh, 2019^{73a}	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	IPD ^b	Yes	Yes	Yes	IPD ^b
Hantsoo, 2017⁷⁴	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Hides, 2007⁷⁵	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Hyphantis, 2011⁷⁶	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	U/C	U/C	Yes	U/C	U/C
Hyphantis, 2014⁷⁷	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Inagaki, 2013⁷⁸	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	No	Yes	Yes	High

Janssen, 2016⁷⁹	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	IPD ^b	Yes	Yes	U/C	U/C
Lamers, 2008⁸⁰	U/C	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	IPD ^b	Yes	Yes	No	U/C
Levin-Aspenson, 2017⁸¹	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Liu, 2016⁸²	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Lotrakul, 2008⁸³	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	No	Yes	Yes	High
Muramatsu, 2007⁸⁴	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Muramatsu, 2018⁸⁵	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Nakku, 2016⁸⁶	No	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	No	Yes	No	High
Paika, 2017⁸⁷	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	U/C	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Park, 2013⁸⁸	U/C	Yes	Yes	U/C	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Persoons, 2001⁸⁹	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Rancans, 2018⁹⁰	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Santos, 2013⁹¹	Yes	Yes	Yes	Low	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	U/C	Yes	Yes	Yes	U/C
Stafford, 2007⁹²	No	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	U/C	U/C
Sung, 2013⁹³	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	Yes	Yes	Low	Low	Yes	Yes	Yes	Yes	Low
Suzuki, 2015⁹⁴	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
van Heyningen, 2018⁹⁵	U/C	Yes	Yes	Low	U/C	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	Yes	Yes	Yes	Yes	Low
Volker, 2016⁹⁶	Yes	Yes	Yes	Low	U/C	N/A	N/A	Low	U/C	Yes	Yes	Yes	Low	U/C	IPD ^b	Yes	Yes	Yes	IPD ^b
Zhang, 2013⁹⁷	U/C	Yes	Yes	U/C	Low	N/A	N/A	Low	Low	Yes	U/C	Yes	U/C	Low	IPD ^b	Yes	Yes	Yes	IPD ^b

Abbreviations: AC: applicability concerns, IPD: individual participant data, N/A: not applicable; RoB: risk of bias, SQ: signalling question, U/C: Unclear

^aWas unpublished at the time of electronic database search

^bRating varies at the individual participant level

SUPPLEMENTARY MATERIAL REFERENCES

1. Amoozegar F, Patten SB, Becker WJ, et al. The prevalence of depression and the accuracy of depression screening tools in migraine patients. *Gen Hosp Psychiatry*. 2017;48:25-31.
2. Amtmann D, Bamer AM, Johnson KL, et al. A comparison of multiple patient reported outcome measures in identifying major depressive disorder in people with multiple sclerosis. *J Psychosom Res*. 2015;79:550-557.
3. Ayalon L, Goldfracht M, Bech P. 'Do you think you suffer from depression?' Re-evaluating the use of a single item question for the screening of depression in older primary care patients. *Int J Geriatr Psychiatry*. 2010;25:497-502.
4. Beraldi A, Baklayan A, Hoster E, Hiddemann W, Heussner P. Which questionnaire is most suitable for the detection of depressive disorders in haemato-oncological patients? Comparison between HADS, CES-D and PHQ-9. *Oncol Res Treat*. 2014;37:108-109.
5. Bernstein CN, Zhang L, Lix LM, et al. The validity and reliability of screening measures for depression and anxiety disorders in inflammatory bowel disease. *Inflamm Bowel Dis*. 2018;24:1867-1875.
6. Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. *BMC Psychiatry*. 2015;15:118.
7. Bombardier CH, Kalpakjian CZ, Graves DE, Dyer JR, Tate DG, Fann JR. Validity of the Patient Health Questionnaire-9 in assessing major depressive disorder during inpatient spinal cord injury rehabilitation. *Arch Phys Med Rehabil*. 2012;93:1838-1845.
8. Chagas MH, Tumas V, Rodrigues GR, et al. Validation and internal consistency of Patient Health Questionnaire-9 for major depression in Parkinson's disease. *Age Ageing*.

2013;42:645-649.

9. Chibanda D, Verhey R, Gibson LJ, et al. Validation of screening tools for depression and anxiety disorders in a primary care population with high HIV prevalence in Zimbabwe. *J Affect Disord.* 2016;198:50-55.
10. Eack SM, Greeno CG, Lee BJ. Limitations of the Patient Health Questionnaire in identifying anxiety and depression in community mental health: Many cases are undetected. *Res Soc Work Pract.* 2006;16:625-631.
11. Fiest KM, Patten SB, Wiebe S, Bulloch AG, Maxwell CJ, Jette N. Validating screening tools for depression in epilepsy. *Epilepsia.* 2014;55:1642-1650.
12. Fischer HF, Klug C, Roeper K, et al. Screening for mental disorders in heart failure patients using computer-adaptive tests. *Qual Life Res.* 2014;23:1609-1618.
13. Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Ann Fam Med.* 2009;7:63-70.
14. Gräfe K, Zipfel S, Herzog W, Löwe B. Screening for psychiatric disorders with the Patient Health Questionnaire (PHQ). Results from the German validation study. *Diagnostica.* 2004;50:171-181.
15. Green JD, Annunziata A, Kleiman SE, et al. Examining the diagnostic utility of the DSM-5 PTSD symptoms among male and female returning veterans. *Depress Anxiety.* 2017;34:752-760.
16. Green EP, Tuli H, Kwobah E, Menya D, Chesire I, Schmidt C. Developing and validating a perinatal depression screening tool in Kenya blending Western criteria with local idioms: a mixed methods study. *J Affect Disord.* 2018;228:49-59.

17. Haroz EE, Bass J, Lee C, et al. Development and cross-cultural testing of the International Depression Symptom Scale (IDSS): a measurement instrument designed to represent global presentations of depression. *GMH*. 2017;4.
18. Hitchon CA, Zhang L, Peschken CA, et al. The validity and reliability of screening measures for depression and anxiety disorders in rheumatoid arthritis. *Arthrit Care Res*. 2019.
19. Khamseh ME, Baradaran HR, Javanbakht A, Mirghorbani M, Yadollahi Z, Malek M. Comparison of the CES-D and PHQ-9 depression scales in people with type 2 diabetes in Tehran, Iran. *BMC Psychiatry*. 2011;11:61.
20. Kwan Y, Tham WY, Ang A. Validity of the Patient Health Questionnaire-9 (PHQ-9) in the screening of post-stroke depression in a multi-ethnic population. *Biol Psychiatry*. 2012;71:141S.
21. Lambert SD, Clover K, Pallant JF, et al. Making sense of variations in prevalence estimates of depression in cancer: A co-calibration of commonly used depression scales using Rasch analysis. *J Natl Compr Canc Netw*. 2015;13:1203-1211.
22. Lara MA, Navarrete L, Nieto L, Martín JP, Navarro JL, Lara-Tapia H. Prevalence and incidence of perinatal depression and depressive symptoms among Mexican women. *J Affect Disord*. 2015;175:18-24.
23. Lino VT, Portela MC, Camacho LA, et al. Screening for depression in low-income elderly patients at the primary care level: use of the Patient Health Questionnaire-2. *PLoS One*. 2014; 9:e113778.
24. Liu SI, Yeh ZT, Huang HC, et al. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Compr Psychiatry*. 2011;52:96-101.
25. Marrie RA, Zhang L, Lix LM, et al. The validity and reliability of screening measures for

- depression and anxiety disorders in multiple sclerosis. *Mult Scler Relat Dis*. 2018;20:9-15.
26. Martin-Subero M, Kroenke K, Diez-Quevedo C, et al. Depression as measured by PHQ-9 versus clinical diagnosis as an independent predictor of long-term mortality in a prospective cohort of medical inpatients. *Psychosom Med*. 2017;79:273-282.
27. McGuire AW, Eastwood JA, Macabasco-O'Connell A, Hays RD, Doering LV. Depression screening: utility of the Patient Health Questionnaire in patients with acute coronary syndrome. *Am J Crit Care*. 2013;22:12-19.
28. Osório FL, Vilela Mendes A, Crippa JA, Loureiro SR. Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context of primary health care. *Perspect Psychiatr Care*. 2009;45:216-227.
29. Osório FL, Carvalho AC, Fracalossi TA, Crippa JA, Loureiro ES. Are two items sufficient to screen for depression within the hospital context? *Int J Psychiatry Med*. 2012;44:141-148.
30. Osório FL, Lima MP, Chagas MH. Screening tools for psychiatry disorders in cancer setting: caution when using. *Psychiatry Res*. 2015;229:739-742.
31. Patten SB, Burton JM, Fiest KM, et al. Validity of four screening scales for major depression in MS. *Mult Scler*. 2015;21:1064-1071.
32. Picardi A, Adler DA, Abeni D, et al. Screening for depressive disorders in patients with skin diseases: a comparison of three screeners. *Acta Derm Venereol*. 2005;85:414-419.
33. Prisdie JC, Fiest KM, Coutts SB, et al. Validating screening tools for depression in stroke and transient ischemic attack patients. *Int J Psychiatry Med*. 2016;51:262-277.
34. Richardson TM, He H, Podgorski C, Tu X, Conwell Y. Screening depression aging services clients. *Am J Geriatr Psychiatry*. 2010;18:1116-1123.
35. Roch S, Fydrich T, Kuech D, et al. Measurement of depression and anxiety in

multidisciplinary inpatient orthopedic rehabilitation - a questionnaire validation with the SCID. *Phys Med Rehab Kuror*. 2016;26:130-136.

36. Rooney AG, McNamara S, Mackinnon M, et al. Screening for major depressive disorder in adults with cerebral glioma: an initial validation of 3 self-report instruments. *Neuro-oncology*. 2013;15:122-129.
37. Shinn EH, Valentine A, Baum G, et al. Comparison of four brief depression screening instruments in ovarian cancer patients: diagnostic accuracy using traditional versus alternative cutpoints. *Gynecol Oncol*. 2017;145:562-568.
38. Sidebottom AC, Harrison PA, Godecker A, Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. *Arch Womens Ment Health*. 2012;15:367-374.
39. Simning A, van Wijngaarden E, Fisher SG, Richardson TM, Conwell Y. Mental healthcare need and service utilization in older adults living in public housing. *Am J Geriatr Psychiatry*. 2012;20:441-451.
40. Spangenberg L, Glaesmer H, Boecker M, Forkmann T. Differences in Patient Health Questionnaire and Aachen Depression Item Bank scores between tablet versus paper-and-pencil administration. *Qual Life Res*. 2015;24:3023-3032.
41. Swartz RH, Cayley ML, Lanctôt KL, et al. The “dOC” screen: feasible and valid screening for depression, obstructive sleep apnea (OSA) and cognitive impairment in stroke prevention clinics. *PLoS One*. 2017;12:e0174451.
42. Turner A, Hambridge J, White J, et al. 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*. 2012;43:1000-1005.
43. Twist K, Stahl D, Amiel SA, Thomas S, Winkley K, Ismail K. Comparison of depressive symptoms in type 2 diabetes using a two-stage survey design. *Psychosom Medicine*. 2013;75:791-797.
 44. Wagner LI, Pugh SL, Small Jr W, et al. Screening for depression in cancer patients receiving radiotherapy: feasibility and identification of effective tools in the NRG Oncology RTOG 0841 trial. *Cancer*. 2017;123:485-493.
 45. Williams JR, Hirsch ES, Anderson K, et al. A comparison of nine scales to detect depression in Parkinson disease: which scale to use? *Neurology*. 2012;78:998-1006.
 46. Wittkamp K, van Ravesteijn H, Baas K, et al. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry*. 2009;31:451-459.
 47. Arroll B, Goodyear-Smith F, Crengle S, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med*. 2010;8:348-353.
 48. Azah MN, Shah ME, Shaaban J, Bahri IS, Rushidi WM, Jamil YM. Validation of the Malay version brief Patient Health Questionnaire (PHQ-9) among adult attending family medicine clinics. *Int Med J Tokyo*. 2005;12:259-263.
 49. De Man-van Ginkel JM, Hafsteinsdóttir T, Lindeman E, Burger H, Grobbee D, Schuurmans M. An efficient way to detect poststroke depression by subsequent administration of a 9-item and a 2-item Patient Health Questionnaire. *Stroke*. 2012;43:854-856.
 50. Delgadillo J, Payne S, Gilbody S, et al. How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. *J Affect Disord*. 2011;134:266-271.

51. Fisher J, Rowe H, Wynter K, et al. Gender-informed, psychoeducational programme for couples to prevent postnatal common mental disorders among primiparous women: cluster randomised controlled trial. *BMJ Open*. 2016;6:e009396.
52. Gelaye B, Tadesse MG, Williams MA, Fann JR, Vander Stoep A, Zhou XH. Assessing validity of a depression screening instrument in the absence of a gold standard. *Ann Epidemiol*. 2014;24:527-531.
53. Grool AM, van der Graaf Y, Mali WP, Geerlings MI. Location of cerebrovascular and degenerative changes, depressive symptoms and cognitive functioning in later life: the SMART-Medea study. *J Neurol Neurosurg Psychiatry*. 2011;82:1093-1100.
54. Hahn D, Reuter K, Harter M. Screening for affective and anxiety disorders in medical patients - comparison of HADS, GHQ-12 and Brief-PHQ. *GMS Psychsoc Med*. 2006;3.
55. Henkel V, Mergl R, Kohnen R, Allgaier AK, Moller HJ, Hegerl U. Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. *Gen Hosp Psychiatry*. 2004;26:190-198.
56. Hobfoll SE, Canetti D, Hall BJ, et al. Are community studies of psychological trauma's impact accurate? A study among Jews and Palestinians. *Psychol Assess*. 2011;23:599-605.
57. Kiely KM, Butterworth P. Validation of four measures of mental health against depression and generalized anxiety in a community based sample. *Psychiatry Res*. 2014;225:291-298.
58. Kim DJ, Kim K, Lee HW, et al. Internet game addiction, depression, and escape from negative emotions in adulthood: a nationwide community sample of Korea. *J Nerv Ment Dis*. 2017;205:568-573.
59. Kohrt BA, Luitel NP, Acharya P, Jordans MJ. Detection of depression in low resource settings: validation of the Patient Health Questionnaire (PHQ-9) and cultural concepts of

- distress in Nepal. *BMC Psychiatry*. 2016;16:58.
60. Liu Y, Wang J. Validity of the patient health questionnaire-9 for DSM-IV major depressive disorder in a sample of Canadian working population. *J Affect Disord*. 2015;187:122-126.
61. Mohd Sidik S, Arroll B, Goodyear-Smith F. Criterion validity of the PHQ-9 (Malay version) in a primary care clinic in Malaysia. *Med J Malaysia*. 2012;67:309-315.
62. Patel V, Araya R, Chowdhary N, et al. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychol Med*. 2008;38:221-228.
63. Pence BW, Gaynes BN, Atashili J, et al. Validity of an interviewer-administered Patient Health Questionnaire-9 to screen for depression in HIV-infected patients in Cameroon. *J Affect Disord*. 2012;143:208-213.
64. Razykov I, Hudson M, Baron M, Thombs BD, Canadian Scleroderma Research Group. Utility of the Patient Health Questionnaire-9 to assess suicide risk in patients with systemic sclerosis. *Arth Care Res*. 2013;65:753-758.
65. Thombs BD, Ziegelstein RC, Whooley MA. Optimizing detection of major depression among patients with coronary artery disease using the Patient Health Questionnaire: data from the heart and soul study. *J Gen Intern Med*. 2008;23:2014-2017.
66. Zuithoff NP, Vergouwe Y, King M, et al. A clinical prediction rule for detecting major depressive disorder in primary care: the PREDICT-NL study. *J Fam Pract*. 2009;26:241-250.
67. Akena D, Joska J, Obuku EA, Stein DJ. Sensitivity and specificity of clinician administered screening instruments in detecting depression among HIV-positive individuals in Uganda. *AIDS Care*. 2013;25:1245-1252.
68. Baron EC, Davies T, Lund C. Validation of the 10-item centre for epidemiological studies

depression scale (CES-D-10) in Zulu, Xhosa and Afrikaans populations in South Africa. *BMC Psychiatry*. 2017;17:6.

69. Buji RI, Abdul Murad NA, Chan LF, et al. Suicidal ideation in systemic lupus erythematosus: NR2A gene polymorphism, clinical and psychosocial factors. *Lupus*. 2018;27:744-752.
70. Cholera R, Gaynes BN, Pence BW, et al. Validity of the Patient Health Questionnaire-9 to screen for depression in a high-HIV burden primary healthcare clinic in Johannesburg, South Africa. *J Affect Disord*. 2014;167:160-166.
71. Conway A, Sheridan J, Maddicks-Law J, et al. Accuracy of anxiety and depression screening tools in heart transplant recipients. *Appl Nurs Res*. 2016;32:177-181.
72. de la Torre AY, Oliva N, Echevarrieta PL, et al. Major depression in hospitalized Argentine general medical patients: prevalence and risk factors. *J Affect Disord*. 2016;197:36-42.
73. Gholizadeh L, Shahmansouri N, Heydari M, Davidson PM. Assessment and detection of depression in patients with coronary artery disease: validation of the Persian version of the PHQ-9. *Contemp Nurse*. 2019;55:185-194.
74. Hantsoo L, Podcasy J, Sammel M, Epperson CN, Kim DR. Pregnancy and the acceptability of computer-based versus traditional mental health treatments. *J Womens Health*. 2017;26:1106-1113.
75. Hides L, Lubman DI, Devlin H, et al. Reliability and validity of the Kessler 10 and Patient Health Questionnaire among injecting drug users. *Aust N Z J Psychiatry*. 2007;41:166-168.
76. Hyphantis T, Kotsis K, Voulgari PV, Tsifetaki N, Creed F, Drosos AA. Diagnostic accuracy, internal consistency, and convergent validity of the Greek version of the Patient Health Questionnaire 9 in diagnosing depression in rheumatologic disorders. *Arthritis Care Res*.

2011;63:1313-1321.

77. Hyphantis T, Kroenke K, Papatheodorou E, et al. Validity of the Greek version of the PHQ 15-item Somatic Symptom Severity Scale in patients with chronic medical conditions and correlations with emergency department use and illness perceptions. *Compr Psychiatry*. 2014;55:1950-1959.
78. Inagaki M, Ohtsuki T, Yonemoto N, et al. Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in general internal medicine primary care at a Japanese rural hospital: a cross-sectional study. *Gen Hosp Psychiatry*. 2013;35:592-597.
79. Janssen EP, Köhler S, Stehouwer CD, et al. The Patient Health Questionnaire-9 as a screening tool for depression in individuals with type 2 diabetes mellitus: the Maastricht study. *J Am Geriatr Soc*. 2016;64:e201-206.
80. Lamers F, Jonkers CC, Bosma H, Penninx BW, Knottnerus JA, van Eijk JT. Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. *J Clin Epidemiol*. 2008;61:679-687.
81. Levin-Aspenson HF, Watson D. Mode of administration effects in psychopathology assessment: analyses of gender, age, and education differences in self-rated versus interview-based depression. *Psychol Assess*. 2018;30:287.
82. Liu H, He YL, Miao JM, Wang JY, Zeng QZ. Validity and reliability of the 12-item Short Form Health Survey in outpatients from traditional Chinese internal department. *Chin Ment Health J*. 2016;30:6-12.
83. Lotrakul M, Sumrithe S, Saipanish R. Reliability and validity of the Thai version of the PHQ-9. *BMC Psychiatry*. 2008;8:46.
84. Muramatsu K, Miyaoka H, Kamijima K, et al. The Patient Health Questionnaire, Japanese

version: validity according to the Mini-International Neuropsychiatric Interview-Plus.
Psychol Rep. 2007;101:952-960.

85. Muramatsu K, Miyaoka H, Kamijima K, et al. Performance of the Japanese version of the Patient Health Questionnaire-9 (J-PHQ-9) for depression in primary care. *Gen Hosp Psychiatry.* 2018;52:64-69.
86. Nakku JE, Rathod SD, Kizza D, et al. Validity and diagnostic accuracy of the Luganda version of the 9-item and 2-item patient health questionnaire for detecting major depressive disorder in rural Uganda. *GMH.* 2016;3.
87. Paika V, Andreoulakis E, Ntountoulaki E, et al. The Greek-Orthodox version of the Brief Religious Coping (B-RCOPE) instrument: psychometric properties in three samples and associations with mental disorders, suicidality, illness perceptions, and quality of life. *Ann Gen Psychiatry.* 2017;16:13.
88. Park H, Kim JH, Hahm BJ. The Distress Thermometer and the PHQ-2 for ultra-brief screening depression of cancer patients in Korea: P3-41. In: Abstracts of the IPOS 15th World Congress of Psycho-Oncology, 4-8 November 2013, Rotterdam, The Netherlands. *Psycho-Oncology.* 2013;22:303-304.
89. Persoons P, Luyckx K, Fischler B. Psychiatric diagnoses in gastroenterology: validation of a self-report instrument (PRIME-MD Patient Health Questionnaire), epidemiology and recognition. *Gastroenterology.* 2001;120:A114.
90. Rancans E, Trapencieris M, Ivanovs R, Vrublevska J. Validity of the PHQ-9 and PHQ-2 to screen for depression in nationwide primary care population in Latvia. *Ann Gen Psychiatry.* 2018;17:33.
91. Santos IS, Tavares BF, Munhoz TN, et al. Sensitivity and specificity of the Patient Health

Questionnaire-9 (PHQ-9) among adults from the general population. *Cad Saude Publica*. 2013;29:1533-1543.

92. Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. *Gen Hosp Psychiatry*. 2007;29:417-424.
93. Sung SC, Low CC, Fung DS, Chan YH. Screening for major and minor depression in a multiethnic sample of Asian primary care patients: a comparison of the nine-item Patient Health Questionnaire (PHQ-9) and the 16-item Quick Inventory of Depressive Symptomatology - Self-Report (QIDS-SR16). *Asia Pac Psychiatry*. 2013;5:249-258.
94. Suzuki K, Kumei S, Ohhira M, Nozu T, Okumura T. Screening for major depressive disorder with the Patient Health Questionnaire (PHQ-9 and PHQ-2) in an outpatient clinic staffed by primary care physicians in Japan: a case control study. *PLoS One*. 2015;10:e0119147.
95. van Heyningen T, Honikman S, Tomlinson M, Field S, Myer L. Comparison of mental health screening tools for detecting antenatal depression and anxiety disorders in South African women. *PLoS One*. 2018;13:e0193697.
96. Volker D, Zijlstra-Vlasveld MC, Brouwers EP, Homans WA, Emons WH, van der Feltz-Cornelis CM. Validation of the patient health questionnaire-9 for major depressive disorder in the occupational health setting. *J Occup Rehabil*. 2016;26:237-244.
97. Zhang Y, Ting R, Lam M, et al. Measuring depressive symptoms using the Patient Health Questionnaire-9 in Hong Kong Chinese subjects with type 2 diabetes. *J Affect Disord*. 2013;151:660-666.
98. Alamri SH, Bari AI, Ali AT. Depression and associated factors in hospitalized elderly: a

- cross-sectional study in a Saudi teaching hospital. *Ann Saudi Med.* 2017;37:122-129.
99. Bailer J, Kerstner T, Witthöft M, Diener C, Mier D, Rist F. Health anxiety and hypochondriasis in the light of DSM-5. *Anxiety Stress Copin.* 2016;29:219-239.
100. Becker S, Al Zaid K, Al Faris E. Screening for somatization and depression in Saudi Arabia: a validation study of the PHQ in primary care. *Int J Psychiatry Med.* 2002;32:271-283.
101. Brodey BB, Goodman SH, Baldasaro RE, et al. Development of the Perinatal Depression Inventory (PDI)-14 using item response theory: a comparison of the BDI-II, EPDS, PDI, and PHQ-9. *Arch Womens Ment Health.* 2016;19:307-316.
102. Chen S, Fang Y, Chiu H, Fan H, Jin T, Conwell Y. Validation of the nine-item Patient Health Questionnaire to screen for major depression in a Chinese primary care population. *Asia Pac Psychiatry.* 2013;5:61-68.
103. Chen S, Conwell Y, Vanorden K, et al. Prevalence and natural course of late-life depression in China primary care: a population based study from an urban community. *J Affect Disord.* 2012;141:86-93.
104. Fann JR, Bombardier CH, Dikmen S, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. *J Head Trauma Rehabil.* 2005;20:501-511.
105. Fisher L, Hessler DM, Polonsky WH, et al. Prevalence of depression in type 1 diabetes and the problem of over-diagnosis. *Diabet Med.* 2016;33:1590-1597.
106. Irmak C, Altıntaş M. The resilience, attachment, coping, and psychopathology of battered women: comparison of sheltered versus in-home women. *Anadolu Psikiyatr De.* 2017;18:561-570.

107. Lai BP, Tang AK, Lee DT, Yip AS, Chung TK. Detecting postnatal depression in Chinese men: a comparison of three instruments. *Psychiatry Res.* 2010;180:80-85.
108. Limon FJ. Screening Latino farmworkers for depression in primary care. *Dissertation Abstracts International: Section B: The Sciences and Engineering.* 2016.
109. Liu ZW, Yu Y, Hu M, Liu HM, Zhou L, Xiao SY. PHQ-9 and PHQ-2 for screening depression in Chinese rural elderly. *PLoS One.* 2016;11:e0151042.
110. Nacak Y, Morawa E, Tuffner D, Erim Y. Insecure attachment style and cumulative traumatic life events in patients with somatoform pain disorder: a cross-sectional study. *J Psychosom Res.* 2017;103:77-82.
111. Navines R, Castellvi P, Moreno-Espana J, et al. Depressive and anxiety disorders in chronic hepatitis C patients: reliability and validity of the Patient Health Questionnaire. *J Affect Disord.* 2012;138:343-351.
112. Phelan E, Williams B, Meeker K, et al. A study of the diagnostic accuracy of the PHQ-9 in primary care elderly. *BMC Fam Pract.* 2010;11:63.
113. Thompson AW, Liu H, Hays RD, et al. Diagnostic accuracy and agreement across three depression assessment measures for Parkinson's disease. *Parkinsonism Relat Disord.* 2011;17:40-45.
114. Vöhringer PA, Jimenez MI, Igor MA, et al. Detecting mood disorder in resource-limited primary care settings: comparison of a self-administered screening tool to general practitioner assessment. *J Med Screen.* 2013;20:118-124.
115. Watnick S, Wang PL, Demadura T, Ganzini L. Validation of 2 depression screening tools in dialysis patients. *Am J Kid Dis.* 2005;46:919-924.
116. Al-Ghafri G, Al-Sinawi H, Al-Muniri A, et al. Prevalence of depressive symptoms as

- elicited by Patient Health Questionnaire (PHQ-9) among medical trainees in Oman. *Asian J Psychiatr.* 2014;8:59-62.
117. Haddad M, Walters P, Phillips R, et al. Detecting depression in patients with coronary heart disease: a diagnostic evaluation of the PHQ-9 and HADS-D in primary care, findings from the UPBEAT-UK study. *PLoS One.* 2013;8:e78493.
118. Ikin JF, McKenzie DP, Gwini SM, et al. Major depression and depressive symptoms in Australian Gulf War veterans 20 years after the Gulf War. *J Affect Disord.* 2016;189:77-84.
119. Smith MV, Gotman N, Lin H, Yonkers KA. Do the PHQ-8 and the PHQ-2 accurately screen for depressive disorders in a sample of pregnant women? *Gen Hosp Psychiatry.* 2010;32:544-548.
120. Valencia-Garcia D, Bi X, Ayón C. Sensitivity and specificity in three measures of depression among Mexican American women. *J Immigr Minor Health.* 2017;19:562-571.
121. Wang L, Lu K, Li J, Sheng L, Ding R, Hu D. Value of patient health questionnaires (PHQ)-9 and PHQ-2 for screening depression disorders in cardiovascular outpatients. *Zhonghua Xin Xue Guan Bing Za Zhi.* 2015;43:428-431.
122. Choi SK, Boyle E, Burchell AN, et al. Validation of six short and ultra-short screening instruments for depression for people living with HIV in Ontario: results from the Ontario HIV treatment network cohort study. *PLoS One.* 2015;10:e0142706.
123. Du N, Yu K, Ye Y, Chen S. Validity study of Patient Health Questionnaire-9 items for Internet screening in depression among Chinese university students. *Asia Pac Psychiatry.* 2017;9:e12266.
124. Griffith SD, Thompson NR, Rathore JS, Jehi LE, Tesar GE, Katzan IL. Incorporating patient-reported outcome measures into the electronic health record for research: application

- using the Patient Health Questionnaire (PHQ-9). *Qual Life Res.* 2015;24:295-303.
125. Margrove K, Mensah S, Thapar A, Kerr M. Depression screening for patients with epilepsy in a primary care setting using the Patient Health Questionnaire-2 and the Neurological Disorders Depression Inventory for Epilepsy. *Epilepsy Behav.* 2011; 21:387-390.
126. Persoons P, Luyckx K, Desloovere C, Vandenberghe J, Fischler B. Anxiety and mood disorders in otorhinolaryngology outpatients presenting with dizziness: validation of the self-administered PRIME-MD Patient Health Questionnaire and epidemiology. *Gen Hosp Psychiatry.* 2003;25:316-323.
127. Rathore JS, Jehi LE, Fan Y, et al. Validation of the Patient Health Questionnaire-9 (PHQ-9) for depression screening in adults with epilepsy. *Epilepsy Behav.* 2014;37:215-220.
128. Scott JD, Wang CC, Coppel E, Lau A, Veitengruber J, Roy-Byrne P. Diagnosis of depression in former injection drug users with chronic hepatitis C. *J Clin Gastroenterol.* 2011;45:462-467.
129. Seo JG, Park SP. Validation of the Patient Health Questionnaire-9 (PHQ-9) and PHQ-2 in patients with migraine. *J Headache Pain.* 2015;16:65.
130. van Steenberg-Weijnenburg KM, de Vroege L, Ploeger RR, et al. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. *BMC Health Serv Res.* 2010;10:235.
131. Wang W, Bian Q, Zhao Y, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. *Gen Hosp Psychiatry.* 2014;36:539-544.
132. Woldetensay YK, Belachew T, Tesfaye M, et al. Validation of the Patient Health Questionnaire (PHQ-9) as a screening tool for depression in pregnant women: Afaan Oromo

version. *PLoS One*. 2018;13:e0191782.

133. Xiong N, Fritzsche K, Wei J, et al. Validation of patient health questionnaire (PHQ) for major depression in Chinese outpatients with multiple somatic symptoms: a multicenter cross-sectional study. *J Affect Disord*. 2015;174:636-643.