Supplementary Material for Fischer et al. “Comparison of different scoring methods based on latent variable models of the PHQ-9: an individual participant data meta-analysis”

Content

[Supplementary Material 1: Search strategy 2](#_Toc55485373)

[Supplementary Material 2: Flowchart 4](#_Toc55485374)

[Supplementary Material 3: CFA Models 5](#_Toc55485375)

[Unidimensional Model 6](#_Toc55485376)

[Twodimensional Model 6](#_Toc55485377)

[Bifactor Model 7](#_Toc55485378)

[Supplementary Material 5: Forest Plots for best cutoffs 9](#_Toc55485379)

[Sum score 9](#_Toc55485380)

[Unidimensional model 10](#_Toc55485381)

[Twodimensional model 11](#_Toc55485382)

[Bifactor model 12](#_Toc55485383)

[Supplementary Material 6: Characteristics of included primary studies in the calibration set (N=24) 13](#_Toc55485384)

[Supplementary Material 7: Characteristics of included primary studies in the validation set (N=17) 15](#_Toc55485385)

[Supplementary Material 8: References of included studies 16](#_Toc55485386)

[Calibration Set 16](#_Toc55485387)

[Validation Set 17](#_Toc55485388)

# Supplementary Material 1: Search strategy

**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. accura\*.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

# Supplementary Material 2: Flowchart



# Supplementary Material 3: CFA Models



## Unidimensional Model

#Loadings

Unidimensional\_Depr\_Factor =~ 1.86055855843524 \* PHQ9\_Q1 + 2.26140308433454 \* PHQ9\_Q2 + 0.992404819520219 \* PHQ9\_Q3 + 1.36127243784208 \* PHQ9\_Q4 + 1.07376446162741 \* PHQ9\_Q5 + 1.62020126739536 \* PHQ9\_Q6 + 1.36136771687569 \* PHQ9\_Q7 + 1.19737609015725 \* PHQ9\_Q8 + 1.42273798856725 \* PHQ9\_Q9

#Thresholds

PHQ9\_Q1 | 0.591081057940429 \*t1 + 2.1222266908974 \*t2 + 2.94383114951583 \*t3

PHQ9\_Q2 | 0.692300810617947 \*t1 + 2.57952911194073 \*t2 + 3.53550920106456 \*t3

PHQ9\_Q3 | -0.04953582978208 \*t1 + 0.869346472581678 \*t2 + 1.36584412948913 \*t3

PHQ9\_Q4 | -0.602895937102372 \*t1 + 0.948903661699967 \*t2 + 1.68625563766773 \*t3

PHQ9\_Q5 | 0.481848534330138 \*t1 + 1.36532703145709 \*t2 + 1.96851497029748 \*t3

PHQ9\_Q6 | 1.15586534208129 \*t1 + 2.24162785122493 \*t2 + 2.83112788871367 \*t3

PHQ9\_Q7 | 0.841483231006097 \*t1 + 1.90288297389649 \*t2 + 2.46296002702646 \*t3

PHQ9\_Q8 | 1.08285982113088 \*t1 + 2.01007362940264 \*t2 + 2.54330972906161 \*t3

PHQ9\_Q9 | 2.17267023376836 \*t1 + 3.09705421774197 \*t2 + 3.56251563702245 \*t3

# Scaling Factors

PHQ9\_Q1 ~\*~ 0.473424665224917\*PHQ9\_Q1

PHQ9\_Q2 ~\*~ 0.404426192683348\*PHQ9\_Q2

PHQ9\_Q3 ~\*~ 0.709797162231844\*PHQ9\_Q3

PHQ9\_Q4 ~\*~ 0.592030860833425\*PHQ9\_Q4

PHQ9\_Q5 ~\*~ 0.681523756193489\*PHQ9\_Q5

PHQ9\_Q6 ~\*~ 0.525221953703416\*PHQ9\_Q6

PHQ9\_Q7 ~\*~ 0.592003947901289\*PHQ9\_Q7

PHQ9\_Q8 ~\*~ 0.641011219847207\*PHQ9\_Q8

PHQ9\_Q9 ~\*~ 0.575037197089645\*PHQ9\_Q9

# Variances

PHQ9\_Q1 ~~ 1\*PHQ9\_Q1

PHQ9\_Q2 ~~ 1\*PHQ9\_Q2

PHQ9\_Q3 ~~ 1\*PHQ9\_Q3

PHQ9\_Q4 ~~ 1\*PHQ9\_Q4

PHQ9\_Q5 ~~ 1\*PHQ9\_Q5

PHQ9\_Q6 ~~ 1\*PHQ9\_Q6

PHQ9\_Q7 ~~ 1\*PHQ9\_Q7

PHQ9\_Q8 ~~ 1\*PHQ9\_Q8

PHQ9\_Q9 ~~ 1\*PHQ9\_Q9

Unidimensional\_Depr\_Factor ~~ 1\*Unidimensional\_Depr\_Factor

## Twodimensional Model

#Loadings

Cognitive\_Affective\_Factor =~ 1.6532381573603 \* PHQ9\_Q1+1.95549935174008 \* PHQ9\_Q2+1.37576887107947 \* PHQ9\_Q6+1.14999904123945 \* PHQ9\_Q7+1.03286614314852 \* PHQ9\_Q8+1.17281156615965 \* PHQ9\_Q9

Somatic\_Factor =~ 0.909403694892871 \* PHQ9\_Q3+1.41686585639893 \* PHQ9\_Q4+1.0251608410275 \* PHQ9\_Q5

#Thresholds

PHQ9\_Q1 | 0.372302637191451 \*t1 + 2.02381646162532 \*t2 + 2.91826670785051 \*t3

PHQ9\_Q2 | 0.41630702225721 \*t1 + 2.41214485588013 \*t2 + 3.44673126438985 \*t3

PHQ9\_Q3 | -0.135886022793312 \*t1 + 0.830321231139953 \*t2 + 1.35389377360073 \*t3

PHQ9\_Q4 | -0.826597258230922 \*t1 + 0.981722281578582 \*t2 + 1.84328066756735 \*t3

PHQ9\_Q5 | 0.435440679245701 \*t1 + 1.39218888448303 \*t2 + 2.05135191440469 \*t3

PHQ9\_Q6 | 0.969245993645282 \*t1 + 2.09536710460082 \*t2 + 2.71408749114442 \*t3

PHQ9\_Q7 | 0.670421821480893 \*t1 + 1.76156975344654 \*t2 + 2.34009384962173 \*t3

PHQ9\_Q8 | 0.963799421741337 \*t1 + 1.92408658378383 \*t2 + 2.49335929691882 \*t3

PHQ9\_Q9 | 2.00362814046883 \*t1 + 2.91976530185346 \*t2 + 3.39889431334211 \*t3

# Scaling Factors

PHQ9\_Q1 ~\*~ 0.517558660906693\*PHQ9\_Q1

PHQ9\_Q2 ~\*~ 0.455299688927812\*PHQ9\_Q2

PHQ9\_Q6 ~\*~ 0.587956641578544\*PHQ9\_Q6

PHQ9\_Q7 ~\*~ 0.656179026436613\*PHQ9\_Q7

PHQ9\_Q8 ~\*~ 0.695583807410991\*PHQ9\_Q8

PHQ9\_Q9 ~\*~ 0.648819171029481\*PHQ9\_Q9

PHQ9\_Q3 ~\*~ 0.73982488273388\*PHQ9\_Q3

PHQ9\_Q4 ~\*~ 0.576629082816523\*PHQ9\_Q4

PHQ9\_Q5 ~\*~ 0.698267711994694\*PHQ9\_Q5

# (Co-)Variances

PHQ9\_Q1 ~~ 1\*PHQ9\_Q1

PHQ9\_Q2 ~~ 1\*PHQ9\_Q2

PHQ9\_Q6 ~~ 1\*PHQ9\_Q6

PHQ9\_Q7 ~~ 1\*PHQ9\_Q7

PHQ9\_Q8 ~~ 1\*PHQ9\_Q8

PHQ9\_Q9 ~~ 1\*PHQ9\_Q9

PHQ9\_Q3 ~~ 1\*PHQ9\_Q3

PHQ9\_Q4 ~~ 1\*PHQ9\_Q4

PHQ9\_Q5 ~~ 1\*PHQ9\_Q5

Cognitive\_Affective\_Factor ~~ 1\*Cognitive\_Affective\_Factor

Somatic\_Factor ~~ 1\*Somatic\_Factor

Cognitive\_Affective\_Factor ~~ 0.888902271230097\*Somatic\_Factor

## Bifactor Model

# Loadings

Cognitive\_Affective\_Factor =~ 0.208074848815783 \* PHQ9\_Q1+1.08428556966816 \* PHQ9\_Q2+0.283725207845012 \* PHQ9\_Q6+-0.287853669721999 \* PHQ9\_Q7+-0.480352072295586 \* PHQ9\_Q8+0.179276373391482 \* PHQ9\_Q9

General\_Factor =~ 1.60430677701327 \* PHQ9\_Q1+2.69438114113628 \* PHQ9\_Q2+0.815525048103322 \* PHQ9\_Q3+1.23680092386828 \* PHQ9\_Q4+0.917079340343974 \* PHQ9\_Q5+1.34744286708542 \* PHQ9\_Q6+1.27170183156061 \* PHQ9\_Q7+1.2706569079012 \* PHQ9\_Q8+1.15750688683762 \* PHQ9\_Q9

Somatic\_Factor =~ 0.448078920271678 \* PHQ9\_Q3+0.583953710462219 \* PHQ9\_Q4+0.476908275737609 \* PHQ9\_Q5

# Thresholds

PHQ9\_Q1 | 0.386091227775555 \*t1 + 2.01248711073149 \*t2 + 2.89397534656308 \*t3

PHQ9\_Q2 | 0.828388599532747 \*t1 + 3.66730585496841 \*t2 + 5.13426973105256 \*t3

PHQ9\_Q3 | -0.138984582828628 \*t1 + 0.838531069408733 \*t2 + 1.36824102461188 \*t3

PHQ9\_Q4 | -0.82225949037227 \*t1 + 0.942975167582264 \*t2 + 1.78489540239349 \*t3

PHQ9\_Q5 | 0.432431381261341 \*t1 + 1.39368086501973 \*t2 + 2.05603056376381 \*t3

PHQ9\_Q6 | 1.01813651878866 \*t1 + 2.14924636636843 \*t2 + 2.76631008272919 \*t3

PHQ9\_Q7 | 0.619014780255909 \*t1 + 1.79279207538633 \*t2 + 2.40989187365669 \*t3

PHQ9\_Q8 | 0.989145095127972 \*t1 + 2.12031450562889 \*t2 + 2.78041882611927 \*t3

PHQ9\_Q9 | 2.0256085184564 \*t1 + 2.94254183447039 \*t2 + 3.42150870047183 \*t3

# Scaling Factors

PHQ9\_Q1 ~\*~ 0.525799319490938\*PHQ9\_Q1

PHQ9\_Q2 ~\*~ 0.325552210845013\*PHQ9\_Q2

PHQ9\_Q3 ~\*~ 0.732084073542339\*PHQ9\_Q3

PHQ9\_Q4 ~\*~ 0.59021157790955\*PHQ9\_Q4

PHQ9\_Q5 ~\*~ 0.695304043134942\*PHQ9\_Q5

PHQ9\_Q6 ~\*~ 0.587615242537674\*PHQ9\_Q6

PHQ9\_Q7 ~\*~ 0.608571008210084\*PHQ9\_Q7

PHQ9\_Q8 ~\*~ 0.592837171997433\*PHQ9\_Q8

PHQ9\_Q9 ~\*~ 0.649301067824203\*PHQ9\_Q9

# (Co-)Variances

General\_Factor ~~ 0\*Cognitive\_Affective\_Factor

General\_Factor ~~ 0\*Somatic\_Factor

Cognitive\_Affective\_Factor ~~ 0\*Somatic\_Factor

General\_Factor ~~ 1\*General\_Factor

Cognitive\_Affective\_Factor ~~ 1\*Cognitive\_Affective\_Factor

Somatic\_Factor ~~ 1\*Somatic\_Factor

PHQ9\_Q1 ~~ 1\*PHQ9\_Q1

PHQ9\_Q2 ~~ 1\*PHQ9\_Q2

PHQ9\_Q3 ~~ 1\*PHQ9\_Q3

PHQ9\_Q4 ~~ 1\*PHQ9\_Q4

PHQ9\_Q5 ~~ 1\*PHQ9\_Q5

PHQ9\_Q6 ~~ 1\*PHQ9\_Q6

PHQ9\_Q7 ~~ 1\*PHQ9\_Q7

PHQ9\_Q8 ~~ 1\*PHQ9\_Q8

PHQ9\_Q9 ~~ 1\*PHQ9\_Q9

# Supplementary Material 5: Forest Plots for best cutoffs

## Sum score



## Unidimensional model


## Twodimensional model



## Bifactor model



# Supplementary Material 6: Sensitivity analysis based on random split

|  |  |  |  |
| --- | --- | --- | --- |
|  | Calibration sample | Validation sample | p-value |
| n | 5100 | 3530 |  |
| Age (mean (SD)) | 50.42 (16.60) | 46.13 (19.30) | <0.001 |
| Male sex (N (%)) | 1985 (39.0) | 1144 (32.4) | <0.001 |
| Country (%) |  | <0.001 |
|  Canada | 484 (9.5) | 777 (22.0) |  |
|  USA | 1290 (25.3) | 903 (25.6) |  |
|  UK | 261 (5.1) | 0 (0.0) |  |
|  Germany | 494 (9.7) | 470 (13.3) |  |
|  Netherlands | 260 (5.1) | 0 (0.0) |  |
|  Australia | 198 (3.9) | 72 (2.0) |  |
|  Brazil | 177 (3.5) | 170 (4.8) |  |
|  Israel | 0 (0.0) | 151 (4.3) |  |
|  Singapore | 0 (0.0) | 113 (3.2) |  |
|  Iran | 122 (2.4) | 0 (0.0) |  |
|  Italy | 0 (0.0) | 138 (3.9) |  |
|  South Africa | 679 (13.3) | 0 (0.0) |  |
|  Mexico | 0 (0.0) | 280 (7.9) |  |
|  Kenya | 0 (0.0) | 192 (5.4) |  |
|  Zimbabwe | 0 (0.0) | 264 (7.5) |  |
|  Spain | 1003 (19.7) | 0 (0.0) |  |
|  Myanmar | 132 (2.6) | 0 (0.0) |  |
| Language (%) |  | <0.001 |
|  English | 2233 (43.8) | 1752 (49.6) |  |
|  German | 494 (9.7) | 470 (13.3) |  |
|  Dutch | 260 (5.1) | 0 (0.0) |  |
|  Portuguese | 177 (3.5) | 170 (4.8) |  |
|  Hebrew | 0 (0.0) | 151 (4.3) |  |
|  Italian | 0 (0.0) | 138 (3.9) |  |
|  Farsi | 122 (2.4) | 0 (0.0) |  |
|  South African languages | 679 (13.3) | 0 (0.0) |  |
|  Spanish | 1003 (19.7) | 280 (7.9) |  |
|  Malay, Chinese or Tamil | 0 (0.0) | 113 (3.2) |  |
|  Kiswahili | 0 (0.0) | 192 (5.4) |  |
|  Shona | 0 (0.0) | 264 (7.5) |  |
|  Burmese | 132 (2.6) | 0 (0.0) |  |
| Method of PHQ-9 administration (N (%)) |  | <0.001 |
|  Face to face | 1528 (30.0) | 1627 (46.1) |  |
|  Internet | 374 (7.3) | 0 (0.0) |  |
|  Self-administered (mail) | 424 (8.3) | 613 (17.4) |  |
|  Self-administered (in research setting) | 2774 (54.4) | 1290 (36.5) |  |
| Method of SCID administration (N (%)) | <0.001 |
|  Face to face | 4175 (81.9) | 2482 (70.3) |  |
|  Computerized (no interviewer) | 147 (2.9) | 0 (0.0) |  |
|  Phone | 778 (15.3) | 1048 (29.7) |  |
| Participant recruitment setting (N (%)) |  | <0.001 |
|  Inpatient care | 1422 (27.9) | 473 (13.4) |  |
|  Non-medical setting | 377 (7.4) | 190 (5.4) |  |
|  Outpatient care | 1634 (32.0) | 2050 (58.1) |  |
|  Primary Care | 1667 (32.7) | 817 (23.1) |  |
| SCID major Depression = yes (N (%)) | 760 (14.9) | 460 (13.0) | 0.015 |
| PHQ-9 total score (mean (SD)) | 6.90 (5.93) | 6.72 (5.96) | 0.156 |

Characteristics of the included participants stratified by sample. For categorical variables, chi-square tests were performed, for continuous variables independent t-tests. M = mean, SD = standard deviation, N = sample size

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Outcome  | Threshold | TP  | FP  | TN  | FN  | Pooled Sensitivity  | Pooled Specificity  | Combined Sensitivity Specificity | τ2 (Sensitivity) | τ2 (Specificity) | R (Sensitivity) | R (Specificity) |
| Sum score | 11  | 587  | 645.6  | 4341.3  | 190.3  | 0.80 [0.68 - 0.88]  | 0.88 [0.84 - 0.92]  | 1.68  | 1.61  | 0.61  | 3.70  | 4.28  |
| Single factor from unidimensional model | 0.59  | 611  | 782.7  | 4204.2  | 166.3  | 0.84 [0.74 - 0.91]  | 0.85 [0.81 - 0.89]  | 1.69  | 1.68  | 0.45  | 3.65  | 3.98  |
| Cognitive-affective factor from two-dimensional model | 0.58  | 615  | 825.2  | 4161.7  | 162.3  | 0.85 [0.75 - 0.92]  | 0.84 [0.80 - 0.88]  | 1.69  | 1.90  | 0.41  | 3.82  | 3.84  |
| General factor from bi-factor model  | 0.56  | 621  | 860.3  | 4126.6  | 156.3  | 0.86 [0.76 - 0.92]  | 0.84 [0.80 - 0.87]  | 1.70  | 1.83  | 0.34  | 3.72  | 3.60  |

Estimates from the IPD meta-analyses for each model’s cut-off maximizing combined sensitivity and specificity, TP = true positives, FP = false positives, TN = true negatives, FN = false negatives, τ2 = tau squared, fractions in TP, FP, TN, FN introduced through statistical weighting of the primary data

#

ROC Curves comparing diagnostic accuracy of the sum score and the latent variable models in the calibration and validation sample

# Supplementary Material 7: Characteristics of included primary studies in the calibration set (N=24)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| First Author, Year | Country | Recruited Population | Diagnostic Interview | Classification System | Total N | Major Depression N (%) |
| Amoozegar, 20171a | Canada | Migraine patients  | SCID | DSM-IV | 203 | 49 (24) |
| Ayalon, 20102 | Israel | Elderly primary care patients | SCID | DSM-IV | 151 | 6 (4) |
| Beraldi, 20143 | Germany | Cancer inpatients | SCID | DSM-IV | 116 | 7 (7) |
| Bombardier, 20124 | USA | Inpatients with spinal cord injuries | SCID | DSM-IV | 160 | 14 (9) |
| Chagas, 20135 | Brazil | Outpatients with Parkinson's Disease | SCID | DSM-IV | 84 | 19 (23) |
| Eack, 20066 | USA | Women seeking psychiatric services for their children at two mental health centers | SCID | DSM-IV | 48 | 12 (27) |
| Fiest, 20147 | Canada | Epilepsy outpatients | SCID | DSM-IV | 169 | 23 (14) |
| Fischer, 20148 | Germany | Heart failure patients | SCID | DSM-IV | 194 | 11 (6) |
| Gjerdingen, 20099 | USA | Mothers registering their newborns for well-child visits at medical or pediatric clinics | SCID | DSM-IV | 419 | 19 (4) |
| Gräfe, 200410 | Germany | Medical and psychosomatic outpatients  | SCID | DSM-IV | 494 | 67 (14) |
| Khamseh, 201111 | Iran | Type 2 diabetes patients | SCID | DSM-IV | 122 | 47 (39) |
| Kwan, 201212 | Singapore | Post-stroke inpatients undergoing rehabilitation | SCID | DSM-IV-TR | 113 | 3 (3) |
| Lambert, 201513a | Australia | Cancer patients | SCID | DSM-IV | 147 | 21 (15) |
| Osório, 200914 | Brazil | Women in primary care | SCID | DSM-IV | 177 | 60 (34) |
| Osório, 201215 | Brazil | Inpatients from various clinical wards | SCID | DSM-IV | 86 | 28 (33) |
| Picardi, 200516 | Italy | Inpatients with skin diseases | SCID | DSM-IV | 138 | 12 (9) |
| Richardson, 201017 | USA | Older adults undergoing in-home aging services care management assessment  | SCID | DSM-IV | 377 | 95 (25) |
| Rooney, 201318 | UK | Patients with cerebral glioma | SCID | DSM-IV | 126 | 14 (12) |
| Sidebottom, 201219 | USA | Pregnant women | SCID | DSM-IV | 246 | 12 (5) |
| Simning, 201220 | USA | Older adults living in public housing | SCID | DSM-IV | 190 | 10 (5) |
| Turner, 201221 | Australia | Stroke patients  | SCID | DSM-IV | 72 | 13 (18) |
| Turner, Unpublishedab | Australia | Cardiac rehabilitation patients | SCID | DSM-IV | 51 | 4 (8) |
| Williams, 201222 | USA | Parkinson’s Disease patients  | SCID | DSM-IV | 235 | 61 (27) |
| Wittkampf, 200923 | The Netherlands | Primary care patients at risk for depression | SCID  | DSM-IV | 260 | 45 (17) |

a Studies that did not come up in our search

b Unpublished studies at time of the electronic search

**Abbreviations**: SCID: the Structured Clinical Interview for DSM; DSM: Diagnostic and Statistical Manual of Mental Disorders; UK: United Kingdom; USA: United States of America.

# Supplementary Material 8: Characteristics of included primary studies in the validation set (N=17)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| First Author, Year | Country | Recruited Population | Diagnostic Interview | Classification System | Total N | Major Depression N(%) |
| Amtmann, 201524 | USA | Multiple sclerosis patients | SCID | DSM-IV | 164 | 48 (29) |
| Bernstein, 201825a | Canada | IBD patients | SCID | DSM-IV | 240 | 21 (9) |
| Bhana, 201526 | South Africa | Chronic care patients | SCID | DSM-IV | 679 | 78 (11) |
| Chibanda, 201627 | Zimbabwe | A primary care population with high HIV prevalence | SCID | DSM-IV | 264 | 149 (56) |
| Green, 201728 | USA  | Returning veterans  | SCID | DSM-V | 176 | 22 (13) |
| Green, 201829 | Kenya | Pregnant women and new mothers | SCID | DSM-V | 192 | 10 (5) |
| Haroz, 201730 | Myanmar | Primary care patients  | SCID | DSM-IV | 132 | 29 (22) |
| Hitchon, 201931a | Canada | Rheumatoid arthritis patients | SCID | DSM-IV | 148 | 16 (11) |
| Lara, 201532 | Mexico | Pregnant women during the third trimester of pregnancy | SCID | DSM-IV | 280 | 29 (10) |
| Marrie, 201833 | Canada | Multiple sclerosis patients | SCID | DSM-IV | 244 | 25 (10) |
| Martin-Subero, 201734 | Spain | Medical inpatients | SCID | DSM-III | 1003 | 83 (8) |
| Patten, 201535 | Canada | Multiple sclerosis patients | SCID | DSM-IV | 143 | 20 (14) |
| Prisnie, 201636 | Canada | Stroke and transient ischemic attack patients | SCID | DSM-IV | 114 | 11 (10) |
| Quinn, Unpublishedab | UK | Stroke patients | SCID | DSM-V | 135 | 15 (12) |
| Shinn, 201737 | USA | Cancer patients | SCID | DSM-IV | 124 | 5 (5) |
| Spangenberg, 201538 | Germany | Primary care patients | SCID | DSM-IV | 160 | 1 (1) |
| Wagner, 201739 | USA | Patients starting radiotherapy for the first diagnosis of any tumor | SCID | DSM-IV | 54 | 6 (11) |

a Studies that did not come up in our search

b Unpublished studies at time of the electronic search

**Abbreviations**: SCID: the Structured Clinical Interview for DSM; DSM: Diagnostic and Statistical Manual of Mental Disorders; UK: United Kingdom; USA: United States of America.

# Supplementary Material 9: References of included studies

## Calibration Set

Please note: We have 23 references instead of 24 because “Turner, unpublished” doesn’t have a citation yet.

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. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. Fiest KM, Patten SB, Wiebe S, Bulloch AG, Maxwell CJ, Jette N. Validating screening tools for depression in epilepsy. Epilepsia. 2014;55:1642–1650.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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–141S.
13. 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.
14. 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.
15. 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.
16. 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.
17. Richardson TM, He H, Podgorski C, Tu X, Conwell Y. Screening depression aging services clients. Am J Geriatr Psychiatry. 2010;18:1116–1123.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. Wittkampf 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.

## Validation Set

Please note: These are 16 references instead of 17 because “Quinn, unpublished” doesn’t have a citation yet.

1. 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.
2. Bernstein CN, Zhang L, Lix LM, Graff LA, 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Patten SB, Burton JM, Fiest KM, et al. Validity of four screening scales for major depression in MS. Mult. Scler. 2015;21:1064-1071.
13. Prisnie JC, Fiest KM, Coutts SB, et al. Validating screening tools for depression in stroke and transient ischemic attack patients. The International Journal of Psychiatry in Medicine. 2016;51:262-277.
14. 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. Gynecologic oncology. 2017;145:562-568.
15. 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.
16. 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;123:485-93.