Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Gen Hosp Psychiatry. Author manuscript; available in PMC 2022 May 17.

Published in final edited form as:

Gen Hosp Psychiatry. 2021 ; 68: 74–82. doi:10.1016/j.genhosppsych.2020.12.007.

Screening for Perinatal Depression with the Patient Health Questionnaire Depression Scale (PHQ-9): A Systematic Review and Meta-analysis

Larry Wang[a], **Kurt Kroenke, MD**[a,b], **Timothy E. Stump, MS**[c], **Patrick O. Monahan, PhD**[c]

[a]Indiana University School of Medicine, Indianapolis, IN, United States.

[b]Regenstrief Institute, Inc., Indianapolis, IN, United States.

[c]Department of Biostatistics, Indiana University Fairbanks School of Public Health and School of Medicine, Indianapolis, IN, USA

Abstract

Objectives: Perinatal depression (PND) is a prevalent and disabling problem both during pregnancy and the postpartum period. The legacy screening measure has been the Edinburgh Postnatal Depression Scale (EPDS). This systematic review examines the validity of the PHQ-9 as a screener for PND.

Methods: The following databases were searched from January 2001 (when the PHQ-9 was first published) through June 2020: MEDLINE, Embase, and PsychInfo. Studies that compared the PHQ-9 to a criterion standard psychiatric interview were used to determine the operating characteristics of sensitivity, specificity and area under the curve (AUC). Studies comparing the PHQ-9 to the EPDS and other depression scales evaluated convergent validity.

Results: A total of 35 articles were eligible for criterion (n=10) or convergent (n=25) validity. Meta-analysis of the 7 criterion validity studies using the standard PHQ-9 cut point ≥ 10 showed a pooled sensitivity, specificity and AUC of 0.84, 0.81 and 0.89, respectively. Operating characteristics of the PHQ-9 and EPDS were nearly identical in head-to-head comparison studies. The median correlation between the PHQ-9 and EPDS was 0.59, and categorical agreement was moderate.

*Corresponding author: Kurt Kroenke, MD, Regenstrief Institute, Rm 221, 1101 W. 10th Street, Indianapolis, IN 46202. Ph 317-274-9046 FAX 317-274-9304. kkroenke@regenstrief.org.

Author Statement

Larry Wang contributed in the conceptualization, data curation, funding acquisition, investigation, methodology, project administration, writing the original draft, and critical review and revision of the manuscript

Kurt Kroenke, MD, contributed in the conceptualization, data curation, investigation, methodology, project administration, supervision, writing the original draft, and critical review and revision of the manuscript

Timothy Stump, MS, contributed in the formal analysis, methodology, validation, visualization, and critical review and revision of the manuscript

Patrick Monahan, PhD, contributed in the formal analysis, supervision, methodology, validation, visualization, and critical review and revision of the manuscript

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Conclusions: The PHQ-9 appears to be a viable option for perinatal depression screening with operating characteristics similar to the legacy EPDS.

Keywords

depression; screening; pregnancy; postpartum; perinatal; PHQ-9

1. Introduction

1.1. Rationale

Perinatal depression (i.e., depression in women during pregnancy or in the postnatal period up to 12 months postpartum) occurs in 10–20% of women [1]. Untreated depression is associated with adverse fetal and newborn outcomes in addition to long-term effects on the mother, child, and family [2]. Numerous guidelines advocate universal perinatal depression screening [2–8]. The legacy screener most commonly recommended and for which there is the greatest amount of evidence is the Edinburgh Postnatal Depression Scale (EPDS). A "legacy" measure is one that is well-validated, widely-used, and considered by many experts to be the standard against which competing measures should be compared [9–11]. The EPDS qualifies as a legacy measure for several reasons. First, it has the largest number of validation studies of any perinatal depression screening measure as summarized in several systematic reviews [12–16]. Second, it is the screening measure most commonly-recommended in perinatal depression guidelines [3, 4]. Third, the 10-item EPDS is brief and has been translated into more than 50 languages [3].

However, many experts also consider the Patient Health Questionnaire depression scale (PHQ-9) as an alternative to the EPDS for perinatal depression screening [2, 5–8]. The PHQ-9 is the most widely used depression measure globally [17] and has been validated across a wide range of age groups, medical conditions, and clinical settings. Since depression is often a chronic or recurrent condition, using a single measure that can assess depression both during and outside the perinatal period may be advantageous in monitoring scores over the lifespan of a woman. The widespread incorporation of the PHQ-9 into healthcare systems, electronic records, and depression screening guidelines makes it a highly familiar metric to clinicians in both primary care and multiple specialty settings [18].

1.2. Objectives

Unlike the EPDS, there has not been a comprehensive assessment of the published literature regarding the validity of the PHQ-9 in screening for perinatal depression. Therefore, we conducted a systematic review and meta-analysis with three objectives:

1. To determine the criterion validity of the PHQ-9 in perinatal depression screening when compared to a criterion standard psychiatric interview;
2. To examine the convergent validity of the PHQ-9 when compared to other validated depression measures in a perinatal population;
3. To compare the performance of the PHQ-9 and EPDS when used in the same studies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Methods

2.1. Identification of studies:

The following databases were searched from January 2001 through June 2020: MEDLINE, Embase, and PsychInfo via PubMed, Embase, and EBSCO search engines respectively. Literature since 2001 was searched since that was the year the first paper on the PHQ-9 was published. The search was formatted to a PICO question with a perinatal population, the PHQ-9 as the intervention, a criterion standard psychiatric interview as the comparator, and depression as the outcome. The search consisted of puerperal disorders or puerperal or postpartum or post-partum or pregnan* or post natal or postnatal or perinatal depression or postpartum depression AND PHQ or patient health questionnaire or patient health questionnaire 9 or PHQ-9 or PHQ 9 AND depress*. The asterisk indicates the search term was truncated and will include variations. For example, “depress*” includes “depression” as well as “depressed” and “depressive”.

2.2. Study selection

Results of the literature search were imported to EndNote X9 where duplicates were removed. Articles were initially screened by one author for relevance to the PHQ-9 predominantly through reading the abstract. If there was insufficient information from the abstract to include or exclude, the full text paper was analyzed. Conference abstracts were not included. The 89 resulting full text articles were independently assessed by both authors to reduce selection bias. Included studies were further sorted into studies that investigated criterion validity or studies that investigated convergent validity. The references of the articles were checked for relevant studies. Inclusion criteria required that the study: 1) utilized the PHQ-9 to assess perinatal depression; 2) assessed either criterion validity of the PHQ-9 using a structured psychiatric interview or its convergent validity by comparing to another validated depression measure; and 3) was published in English or had an English translation available.

3.1. Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool uses 11 questions to assess the risk of bias of four domains-- patient selection (3 questions), index test (2 questions), reference standard (2 questions), and flow and timing (4 questions).[19] The tool has three choices for each question -- yes, no, and unclear. An answer of “yes” to a question indicates a fulfillment of the QUADAS-2 criteria. An answer of “no” to a question indicates a risk of bias in that domain. An answer of “unclear” indicates the study only partially fulfills the domain or there is insufficient information to draw a conclusion. To be conservative, an answer of “unclear” is treated as a “no”. If each domain contains all “yes” answers, the risk of bias is considered to be low. Developers of the QUADAS consider one or more “no” answers for a domain to indicate a high risk of bias for that domain. Because this seems particularly conservative, we adapted the scoring so that one “no” answer was considered to represent a medium risk of bias, whereas two or more “no” answers was considered to represent a high risk. Two authors appraised each of the articles independently and blind to the other’s results. Consensus was achieved by discussing any ratings that differed between the two reviewers.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Two of the 11 questions required operationalization for this review. Under the index test bias domain, one question is whether the "index test results were interpreted without knowledge of the results of the reference standard." While blinded scoring of the PHQ-9 was not explicitly noted in any of the studies, this is irrelevant since the PHQ-9 is a self-administered scale for which knowing the reference results cannot influence its scoring. Under the flow and timing bias domain, one question is whether "there was an appropriate interval between the index test and the reference standard." A two-week interval or less was deemed appropriate since this is the standard duration DSM uses for defining clinically relevant depressive symptoms.

3.4. Data Extraction and Analysis

Data extracted for each study included sample size, mean age, country and clinical setting in which the study was conducted, perinatal population (pregnancy, postpartum, or both), and proportion of participants with a PHQ-9 score ≥ 10. For Objective 1, diagnostic operating characteristics for criterion validity studies included sensitivity (percent of patients with major depression who have a depression screener score at or above the defined cutpoint), specificity (percent of patients without major depression who have a depression screener score below the defined cutpoint), and area under the curve (AUC) as determined by receiver operating characteristic (ROC) curve analyses. AUC represents the true positive rate divided by the false positive rate across a series of cutpoints. AUCs ≥ 0.80 and ≥ 0.90 represent good and excellent diagnostic test accuracy, respectively.

For studies that used the conventional PHQ-9 threshold score ≥ 10, two-by-two tables were generated from the presented data. From these tables, we calculated the sensitivity, specificity and confidence intervals (CIs) for each study included. The data were presented in the form of forest plots. For all summary-level estimates, we used a bivariate generalized linear mixed model to simultaneously estimate pooled measures of sensitivity and specificity while accounting for the potential correlation between sensitivity and specificity [20]. Summary receiver-operating characteristic curves were obtained along with 95% confidence regions for the bivariate estimate of AUC. The glmer function [21] of the lme4 package [22] in R (R Foundation for Statistical Computing) [23] was used to estimate the bivariate models.

For Objective 2, studies that compared the PHQ-9 to another validated depression measure were summarized to examine convergent validity. Objective 3 focused on the subset of criterion or convergent validity studies that administered the PHQ-9 and the legacy EPDS depression scale to the same patient sample; for this objective, psychometric characteristics of the two scales were compared, including diagnostic operating characteristics, correlation, categorical agreement, and other metrics reported for both scales.

3. Results

As summarized in the PRISMA diagram (Figure 1), 2229 articles were identified through the literature search, of which 1675 abstracts were screened once duplicates were removed. Of 89 full-text articles reviewed after abstract screening, 35 articles were eligible, including 10 for criterion validity [24–33] and 25 for convergent validity [34–58].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3.1. Study characteristics

Table 1 summarizes the 35 studies included in this review. A total of 19,760 women were included in these studies, with a median sample size of 293 (range, 56 to 3342). There were 15 studies that included pregnant women only, 13 that focused only on the postpartum period, and 7 studies that included both pregnant women and those in the postpartum period. Thirteen studies were conducted in the United States, 9 in Africa, 4 in Europe, 4 in Asia, 4 in Latin America, and 1 in Australia. The mean age of women in the studies was a median of 28.2 years (range, 18 to 34.6). Seven of the studies were conducted in an antenatal clinic; 5 in a hospital; 3 each in an obstetrics clinic, a psychiatry setting, the community, a clinical trial, or by a phone survey; 2 each in a postpartum clinic, a family medicine clinic, or a cohort study; and 1 each in pediatrics and an HIV clinic. Of the 19 studies that reported the proportion of the sample having clinically relevant depressive symptoms as noted by the conventional PHQ-9 cutpoint ≥ 10, the median proportion was 18% (range, 5% to 100%).

3.2. Criterion validity

Table S1 summarizes the 10 studies that examined criterion validity using a structured psychiatric interview. A total of 5,235 women were included in these 10 studies, with a median sample size of 311 (range, 160 to 1731). The most common criterion standard interviews were the SCID (n = 5) and MINI (n = 2). In the 9 studies reporting the prevalence of major depression, the median was 11.3% (range, 3.6% to 72.4%).

Figure 2 shows the forest plots for the 7 studies that used the conventional PHQ-9 threshold score ≥ 10. The pooled sensitivity was 0.84 (95% CI, 0.75 to 0.90) and the pooled specificity was 0.81 (95% CI, 0.74 to 0.86). Of note, the 95% CIs were reasonably narrow suggesting relatively precise point estimates. Figure S1 shows a pooled AUC of 0.89 with a reasonably precise elliptical confidence region. In 3 studies that also examined the PHQ-2, a cutpoint of 2 or 3 on the latter had relatively comparable operating characteristics to the PHQ-9.

3.3. Quality assessment

Table S2 summarizes the QUADAS ratings for the 10 studies that examined criterion validity. A low risk of bias in all 4 domains was achieved by 1 study, in 3 domains by 4 studies, in 2 domains by 2 studies, and in only 1 domain by 3 studies. The study by Sidebottom et al. had the least bias and the second largest sample size (n = 745, or 19% of the total patients in the 7 studies using a PHQ-9 ≥ 10 cutpoint).[27] Moreover, operating characteristics in the Sidebottom et al study were higher than the calculated median. The highest risk of bias occurred in the reference standard domain, with only 3 of the 10 criterion validity studies having the lowest risk of bias. The risk was due to six articles being unclear if blinding occurred when giving the reference standard, with only one article[28] explicitly stating a lack of blinding. For the 110 QUADAS questions rated (11 questions across 10 articles), the two raters agreed 85% of the time for a kappa of 0.47. Disagreements were discussed in order to achieve a consensus on final QUADAS ratings.

3.4. Psychometric comparison of PHQ-9 and EPDS

Of 19 studies administering both the PHQ-9 and EPDS to participants, 15 provided one or more psychometric comparisons between the two scales (Table 2). The median sensitivity

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of the PHQ-9 and EPDS were remarkably similar (.81 vs .82 in 5 studies) as were the median specificity (.75 vs .73 in 5 studies) and the median AUC (.86 vs .88 in 5 studies). In 6 studies, the median correlation between the two scales was 0.59. In 4 studies, there was moderate categorical agreement as assessed by simple agreement or kappa (agreement beyond chance). Responsiveness and the rates of moderate depression were examined in two studies each and were similar using the PHQ-9 and EPDS. Finally, test-retest reliability was compared in only 1 study and was somewhat higher for the PHQ-9 (0.75 vs. 0.51).

4. Discussion

Our systematic review has several major findings. First, the PHQ-9 has good diagnostic operating characteristics as a screener for perinatal depression; its sensitivity, specificity, and AUC all > 0.80 is similar to the performance of other well-validated depression measures used across a range of clinical conditions [59–63]. Second, the PHQ-9 appears to perform comparably to the EPDS which heretofore has been the legacy perinatal depression scale. Third, the wide diversity of study samples, countries and clinical settings in which the PHQ-9 has been examined for perinatal depression assessment (Table 1) increases the potential generalizability of our study results to the various settings in which depression screening may be warranted. The widespread use of the PHQ-9 coupled with recommendations for universal perinatal depression screening enhances the importance of our findings.

Notably, the pooled sensitivity and specificity of 0.84 and 0.81 for a PHQ-9 score ≥ 10 from our meta-analysis is very similar to the 0.85 and 0.84 reported for a standard EPDS cutpoint score ≥ 10 in a recent meta-analysis [16]. Paralleling these findings, the studies in our systematic review that compared the PHQ-9 and EPDS in the same patient sample found the two measures to have a comparable median sensitivity, specificity, and AUC (Table 2). Both correlations as well as categorical agreement between the PHQ-9 and EPDS were moderate. However, even agreement between criterion standard interviews can vary substantially [64, 65]. Thus, the comparable results for the PHQ-9 and EPDS in Table 2 suggest that either scale may be a reasonable option for perinatal depression screening. Both the EPDS and PHQ-9 are equally brief (10 and 9 items, respectively) and also freely available. This is in contrast to some depression measures which are 20 items or longer and, in some cases, proprietary.

Certainly, the EPDS has had a larger number of validation studies, warranting further criterion validity studies of the PHQ-9 in larger samples as well as additional head-to-head comparisons with the EPDS. That being said, it is also important to comment on three differences between the EPDS and PHQ-9. First, the EPDS does not include the four somatic symptoms of major depressive disorder (fatigue, sleep disturbance, change in weight or appetite, and psychomotor agitation/retardation). This is sometimes cited as an advantage because such symptoms may be common in pregnancy and not specific for depression. Paradoxically, however, validation studies typically compare the EPDS to criterion standard interviews for major depression that include these four somatic symptoms. Importantly, debates about how to handle somatic symptoms in patients with medical or other conditions have tended to favor an inclusive approach because: a) including somatic symptoms is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

more sensitive and reliable [66]; and b) these four somatic symptoms demonstrate robust improvement with treatment, and this improvement does not differ significantly between patients with and without medical co-morbidity [67] Second, the EPDS includes two anxiety symptoms (anxious/worried and scared/panicky) which, although common in perinatal depression, are not part of the diagnostic criteria for depressive disorders. The GAD-7 and the GAD-2 are more specific for anxiety and have been validated for perinatal anxiety screening [68, 69]. Third, the EPDS is a perinatal-specific depression measure whereas the PHQ-9 has the advantage of assessing depression across the lifespan of women [53, 70–72]. Premenstrual and menopausal mood disorders are common in women as well as other chronic depressive disorders; moreover, depression is often a recurrent condition. Thus, using a common measure throughout a woman's life may be preferable to using one measure during the perinatal period and a different measure for assessing depression at other times. Translated into more than 100 languages, the PHQ-9 is public domain and now the most commonly used depression measure in both clinical practice and research [18, 73].

A two-step approach to screening is sometimes used with administration of the PHQ-2 as a first-step screener and, if positive (using a cutpoint of either 2 or 3 on this 6-point scale), completion of the full PHQ-9 [74, 75]. Notably, the three studies in our review that reported on the PHQ-2 showed operating characteristics as a screener that compared favorably with the longer PHQ-9.[24, 25, 30] Further, a yes-no version of the PHQ-2 (instead of the scored version) with or without a follow-up PHQ-9 was found have comparable operating characteristics to the EPDS and greater cost-effectiveness.[76, 77]

Screening for suicidality is an essential step in evaluating depression. Item 9 of the PHQ-9 assesses suicidal ideation over the last 14 days while item 10 of the EPDS assesses suicidal ideation in the past 7 days. A study of 1,517 pregnant women in Peru found that the two suicidal ideation items had a high concordance rate (84.2%) and moderate agreement beyond chance (kappa = 0.42) [78]. Based on the PHQ-9 and the EPDS, 15.8 and 8.8 % of participants screened positive for suicidal ideation, respectively. Responsiveness to treatment (also called sensitivity to change to an anchor or criterion, with or without treatment) is another valuable attribute of a depression measure in clinical settings. The PHQ-9 has proven sensitive to change in prior studies [72, 79–81] and its responsiveness was similar to the EPDS in our review [47, 49].

Anxiety is as prevalent as depression in the perinatal population and the two conditions co-occur 40% of the time. Thus, guidelines frequently recommend joint screening for anxiety and depression [2, 3, 82–85]. The EPDS has two anxiety items and has AUCs of 0.62 to 0.73 in screening for perinatal anxiety [85]. However, the GAD-7 is more commonly recommended if anxiety screening is desired and in one study proved superior to the EPDS [86]. The PHQ-4 combines the PHQ-2 depression scale and GAD-2 anxiety scale and has proven an effective screener in an international study of 1,148 pregnant women.[87]

Prior systematic reviews [4, 15, 88, 89] may have differed in their conclusions regarding depression screeners due to the particular studies included (or criteria for excluding studies), wide variation among studies/samples for findings regarding the same scale, and the smaller number of studies doing head-to-head comparisons of 2 or more scales against a structured

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

psychiatric interview. Also, a recent meta-analysis found that depression prevalence varied widely among perinatal depression studies [90], a finding consistent with the studies included in our systematic review.

One limitation of our findings is that the risk of bias varied among the criterion validity studies, with only 5 of the 10 studies having a low risk of bias in 3 or all 4 of the four QUADAS domains. Also, only 7 of the 10 studies used the standard PHQ-9 ≥ 10 cutpoint. However, the remarkable similarity that we found between the operating characteristics of the PHQ-9 and EPDS is reassuring. Whereas the EPDS has been widely studied for perinatal depression screening, our review is the first to systematically compare its performance with the PHQ-9 in studies administering both scales.

Examples of the increased use of the PHQ-9 during pregnancy and the postpartum period include serving as the criterion standard or primary outcome in clinical trials [91–93] and large cohort studies [94, 95] as well as perinatal screening in large healthcare systems [96, 97]. Interestingly, one study found the PHQ-9 and EPDS comparable in a non-pregnant population which also included men [98]. Also, both scales have been used to screen for antenatal and postnatal paternal depression.[99–102]. Results from ongoing studies comparing the PHQ-9 and EPDS may further inform scale selection [103, 104]. Treatment of perinatal depression has proven effective in multiple trials [105], further highlighting the importance of screening with efficient and evidence-based measures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

This project was funded with support from the Indiana Clinical and Translational Sciences Institute funded, in part, by UL1TR002529 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health

References

[1]. Woody CA, Ferrari AJ, Siskind DJ, Whiteford HA and Harris MG. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. J Affect Disord 2017;219:86–92. [PubMed: 28531848]

[2]. Yeaton-Massey A and Herrero T. Recognizing maternal mental health disorders: beyond postpartum depression. Curr Opin Obstet Gynecol 2019;31:116–119. [PubMed: 30694850]

[3]. ACOG Committee Opinion No. 757 Summary: Screening for Perinatal Depression. Obstet Gynecol 2018;132:1314–1316. [PubMed: 30629563]

[4]. Siu AL. Screening for depression in adults: US Preventive Services Task Force recommendation statement. JAMA 2016;315:380–387. [PubMed: 26813211]

[5]. Earls MF, Yogman MW, Mattson G, Rafferty J, Committee On Psychosocial Aspects Of C and Family H. Incorporating Recognition and Management of Perinatal Depression Into Pediatric Practice. Pediatrics 2019;143. [PubMed: 31060547]

[6]. Learman LA. Screening for Depression in Pregnancy and the Postpartum Period. Clin Obstet Gynecol 2018;61:525–532. [PubMed: 29389681]

[7]. Smith EK, Gopalan P, Glance JB and Azzam PN. Postpartum depression screening: A review for psychiatrists. Harvard Review of Psychiatry 2016;24:173–187. [PubMed: 27148910]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[8]. Stewart DE and Vigod S. Postpartum Depression. N Engl J Med 2016;375:2177–2186. [PubMed: 27959754]

[9]. Gibbons LE, Feldman BJ, Crane HM, Mugavero M, Willig JH, Patrick D, et al. Migrating from a legacy fixed-format measure to CAT administration: calibrating the PHQ-9 to the PROMIS depression measures. Qual Life Res 2011;20:1349–57. [PubMed: 21409516]

[10]. Levin JB, Aebi ME, Smyth KA, Tatsuoka C, Sams J, Scheidemantel T, et al. Comparing Patient-Reported Outcomes Measure Information System Depression Scale with Legacy Depression Measures in a Community Sample of Older Adults with Varying Levels of Cognitive Functioning. Am J Geriatr Psychiatry 2015;23:1134–43. [PubMed: 26320720]

[11]. Nayfe R, Chansard M, Hynan LS, Mortensen EM, Annaswamy T, Fraenkel L, et al. Comparison of patient-reported outcomes measurement information system and legacy instruments in multiple domains among older veterans with chronic back pain. BMC Musculoskelet Disord 2020;21:598. [PubMed: 32900386]

[12]. Kozinszky Z and Dudas RB. Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period. J Affect Disord 2015;176:95–105. [PubMed: 25704562]

[13]. Owora AH, Carabin H, Reese J and Garwe T. Summary diagnostic validity of commonly used maternal major depression disorder case finding instruments in the United States: A meta-analysis. Journal of Affective Disorders 2016;205:335–343. [PubMed: 27566453]

[14]. O'Connor E, Rossom RC, Henninger M, Groom HC and Burda BU. Primary care screening for and treatment of depression in pregnant and postpartum women: Evidence report and systematic review for the US Preventive Services Task Force. JAMA 2016;315:388–406. [PubMed: 26813212]

[15]. Ukatu N, Clare CA and Brulja M. Postpartum Depression Screening Tools: A Review. Psychosomatics 2018;59:211–219. [PubMed: 29396166]

[16]. Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BDand Group DESDE. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. BMJ 2020;371:m4022. [PubMed: 33177069]

[17]. Hirschtritt ME and Kroenke K. Screening for Depression. JAMA 2017;318:745–746. [PubMed: 28829850]

[18]. Kroenke K The PHQ-9: global uptake of a depression scale World Psychiatry in press.

[19]. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36. [PubMed: 22007046]

[20]. Chu H and Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol 2006;59:1331–2; author reply 1332–3. [PubMed: 17098577]

[21]. Partlett C and Takwoingi Y. Meta-analysis of test accuracy studies in R: a summary of user-written programs and step-by-step guide to using glmer. Version 1.0. 2016.

[22]. Bates D, Machler M, Bolker BM and Walker SC. Fitting linear mixed-effects models using lme4. J Stat Software 2015;67:1–48.

[23]. R Core Team. R: a language and environment for statistical computing. Vienna: Foundation for Statistical Computing, 2013.

[24]. Gjerdingen D, Crow S, McGovern P, Miner M and Center B. Postpartum depression screening at well-child visits: Validity of a 2-question screen and the PHQ-9. Annals of Family Medicine 2009;7:63–70. [PubMed: 19139451]

[25]. Smith MV, Gotman N, Lin H and Yonkers KA. Do the PHQ-8 and the PHQ-2 accurately screen for depressive disorders in a sample of pregnant women? General Hospital Psychiatry 2010;32:544–548. [PubMed: 20851275]

[26]. Flynn HA, Sexton M, Ratliff S, Porter K and Zivin K. Comparative performance of the Edinburgh Postnatal Depression Scale and the Patient Health Questionnaire-9 in pregnant and postpartum women seeking psychiatric services. Psychiatry Research 2011;187:130–134. [PubMed: 21122923]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[27]. Sidebottom AC, Harrison PA, Godecker A and Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. Archives of Women's Mental Health 2012;15:367–374.

[28]. Gallis JA, Maselko J, O'Donnell K, Song K, Saqib K, Turner EL, et al. Criterion-related validity and reliability of the Urdu version of the patient health questionnaire in a sample of community-based pregnant women in Pakistan. PeerJ 2018;6:e5185. [PubMed: 30038858]

[29]. Woldetensay YK, Belachew T, Tesfaye M, Spielman K, Biesalski HK, Kantelhardt EJ, 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.

[30]. van Heyningen T, Honikman S, Tomlinson M, Field S and Myer L. Comparison of mental health screening tools for detecting antenatal depression and anxiety disorders in South African women. PLoS ONE 2018;13.

[31]. Weobong B, Akpalu B, Doku V, Owusu-Agyei S, Hurt L, Kirkwood B, et al. The comparative validity of screening scales for postnatal common mental disorder in Kintampo, Ghana. Journal of Affective Disorders 2009;113:109–117. [PubMed: 18614241]

[32]. Green EP, Tuli H, Kwobah E, Menya D, Chesire I and Schmidt C. Developing and validating a perinatal depression screening tool in Kenya blending Western criteria with local idioms: A mixed methods study. Journal of Affective Disorders 2018;228:49–59. [PubMed: 29227955]

[33]. Brodey BB, Goodman SH, Baldasaro RE, Brooks-DeWeese A, Wilson ME, Brodey ISB, 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. Archives of Women's Mental Health 2016;19:307–316.

[34]. Barthel D, Barkmann C, Ehrhardt S, Schoppen S and Bindt C. Screening for depression in pregnant women from Côte d'Ivoire and Ghana: Psychometric properties of the Patient Health Questionnaire-9. Journal of Affective Disorders 2015;187:232–240. [PubMed: 26343851]

[35]. Beck CT, Kurz B and Gable RK. Concordance and discordance of the Postpartum Depression Screening Scale and Patient Health Questionnaire-9 in an ethnically diverse sample. Journal of Social Service Research 2012;38:439–450.

[36]. Buttner MM, Mott SL, Pearlstein T, Stuart S, Zlotnick C and O'Hara MW. Examination of premenstrual symptoms as a risk factor for depression in postpartum women. Archives of Women's Mental Health 2013;16:219–225.

[37]. Davis K, Pearlstein T, Stuart S, O'Hara M and Zlotnick C. Analysis of brief screening tools for the detection of postpartum depression: Comparisons of the PRAMS 6-item instrument, PHQ-9, and structured interviews. Archives of Women's Mental Health 2013;16:271–277.

[38]. Di Venanzio C, Pacitti F, Rossetti MC, Santarelli V, Gregori E, D'Alfonso A, et al. Perinatal depression screening and early treatment. Journal of Psychopathology 2017;23:99–104.

[39]. Gawlik S, Waldeier L, Müller M, Szabo A, Sohn C and Reck C. Subclinical depressive symptoms during pregnancy and birth outcome—A pilot study in a healthy German sample. Archives of Women's Mental Health 2013;16:93–100.

[40]. Gelaye B, Zhong Q-Y, Basu A, Levey EJ, Rondon MB, Sanchez S, et al. Trauma and traumatic stress in a sample of pregnant women. Psychiatry Research 2017;257:506–513. [PubMed: 28843870]

[41]. Hanusa BH, Scholle SH, Haskett RF, Spadaro K and Wisner KL. Screening for depression in the postpartum period: A comparison of three instruments. Journal of Women's Health 2008;17:585–596.

[42]. Harrington BJ, Hosseinipour MC, Maliwichi M, Phulusa J, Jumbe A, Wallie S, et al. Prevalence and incidence of probable perinatal depression among women enrolled in Option B+ antenatal HIV care in Malawi. Journal of Affective Disorders 2018;239:115–122. [PubMed: 29990658]

[43]. Joshi U, Lyngdoh T and Shidhaye R. Validation of Hindi version of Edinburgh Postnatal Depression Scale as a screening tool for antenatal depression. Asian Journal of Psychiatry 2020;48.

[44]. Kadir AA, Daud MNM, Yaacob MJ and Hussain NHN. Relationship between obstetric risk factors and postnatal depression in Malaysian women. International Medical Journal 2009;16:101–106.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[45]. Kulathilaka S, Hanwella R and de Silva VA. Depressive disorder and grief following spontaneous abortion. BMC Psychiatry 2016;16. [PubMed: 26812906]

[46]. Lara MA, Navarrete L, Nieto L, Barba Martín JP, Navarro JL and Lara-Tapia H. Prevalence and incidence of perinatal depression and depressive symptoms among Mexican women. Journal of Affective Disorders 2015;175:18–24. [PubMed: 25590762]

[47]. Loughnan SA, Butler C, Sie AA, Grierson AB, Chen AZ, Hobbs MJ, et al. A randomised controlled trial of 'MUMentum postnatal': Internet-delivered cognitive behavioural therapy for anxiety and depression in postpartum women. Behaviour Research and Therapy 2019;116:94–103. [PubMed: 30877878]

[48]. Maliszewska K, Świątkowska-Freund M, Bidzan M and Preis K. Screening for maternal postpartum depression and associations with personality traits and social support A Polish follow-up study 4 weeks and 3 months after delivery. Psychiatria Polska 2017;51:889–898. [PubMed: 29289968]

[49]. Meltzer-Brody S, Brandon AR, Pearson B, Burns L, Raines C, Bullard E, et al. Evaluating the clinical effectiveness of a specialized perinatal psychiatry inpatient unit. Archives of Women's Mental Health 2014;17:107–113.

[50]. Miller LJ, McGlynn A, Suberlak K, Rubin LH, Miller M and Pirec V. Now what? Effects of on-site assessment on treatment entry after perinatal depression screening. Journal of Women's Health 2012;21:1046–1052.

[51]. Mochache K, Mathai M, Gachuno O, Vander Stoep A and Kumar M. Depression during pregnancy and preterm delivery: a prospective cohort study among women attending antenatal clinic at Pumwani Maternity Hospital. Ann Gen Psychiatry 2018;17:31. [PubMed: 30061917]

[52]. Nieminen K, Berg I, Frankenstein K, Viita L, Larsson K, Persson U, et al. Internet-provided cognitive behaviour therapy of posttraumatic stress symptoms following childbirth —A randomized controlled trial. Cognitive Behaviour Therapy 2016;45:287–306. [PubMed: 27152849]

[53]. Orta OR, Gelaye B, Qiu C, Stoner L and Williams MA. Depression, anxiety and stress among pregnant migraineurs in a pacific-northwest cohort. Journal of Affective Disorders 2015;172:390–396. [PubMed: 25451442]

[54]. Osok J, Kigamwa P, Stoep AV, Huang K-Y and Kumar M. Depression and its psychosocial risk factors in pregnant Kenyan adolescents: A cross-sectional study in a community health Centre of Nairobi. BMC Psychiatry 2018;18. [PubMed: 29357830]

[55]. Sanchez SE, Puente GC, Atencio G, Qiu C, Yanez D, Gelaye B, et al. Risk of spontaneous preterm birth in relation to maternal depressive, anxiety, and stress symptoms. J Reprod Med 2013;58:25–33. [PubMed: 23447915]

[56]. Sefogah PE, Samba A, Mumuni K and Kudzi W. Prevalence and key predictors of perinatal depression among postpartum women in Ghana. Int J Gynaecol Obstet 2020;149:203–210. [PubMed: 32078159]

[57]. Yawn BP, Pace W, Wollan PC, Bertram S, Kurland M, Graham D, et al. Concordance of Edinburgh Postnatal Depression Scale (EPDS) and Patient Health Questionnaire (PHQ-9) to assess increased risk of depression among postpartum women. J Am Board Fam Med 2009;22:483–91. [PubMed: 19734393]

[58]. Zhong Q, Gelaye B, Rondon M, Sánchez SE, García PJ, Sánchez E, et al. Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression. Journal of Affective Disorders 2014;162:1–7. [PubMed: 24766996]

[59]. von Glischinski M, von Brachel R and Hirschfeld G. How depressed is "depressed"? A systematic review and diagnostic meta-analysis of optimal cut points for the Beck Depression Inventory revised (BDI-II). Qual Life Res 2019;28:1111–1118. [PubMed: 30456716]

[60]. Vilagut G, Forero CG, Barbaglia G and Alonso J. Screening for Depression in the General Population with the Center for Epidemiologic Studies Depression (CES-D): A Systematic Review with Meta-Analysis. PLoS One 2016;11:e0155431.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[61]. Levis B, Benedetti A, Thombs BDand Collaboration DESD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019;365:l1476. [PubMed: 30967483]

[62]. Vodermaier A and Millman RD. Accuracy of the Hospital Anxiety and Depression Scale as a screening tool in cancer patients: a systematic review and meta-analysis. Support Care Cancer 2011;19:1899–908. [PubMed: 21898134]

[63]. Krishnamoorthy Y, Rajaa S and Rehman T. Diagnostic accuracy of various forms of geriatric depression scale for screening of depression among older adults: Systematic review and meta-analysis. Arch Gerontol Geriatr 2020;87:104002. [PubMed: 31881393]

[64]. Gjerdingen D, McGovern P and Center B. Problems with a diagnostic depression interview in a postpartum depression trial. J Am Board Fam Med 2011;24:187–93. [PubMed: 21383219]

[65]. Levis B, McMillan D, Sun Y, He C, Rice DB, Krishnan A, et al. Comparison of major depression diagnostic classification probability using the SCID, CIDI, and MINI diagnostic interviews among women in pregnancy or postpartum: An individual participant data meta-analysis. Int J Methods Psychiatr Res 2019;28:e1803. [PubMed: 31568624]

[66]. Koenig HG, George LK, Peterson BL and Pieper CF. Depression in medically ill hospitalized older adults: prevalence, characteristics, and course of symptoms according to six diagnostic schemes. Am J Psychiatry 1997;154:1376–83. [PubMed: 9326819]

[67]. Simon GE and Von Korff M. Medical co-morbidity and validity of DSM-IV depression criteria. Psychol Med 2006;36:27–36. [PubMed: 16202189]

[68]. Nelson HD, Cantor A, Pappas M and Weeks C. Screening for Anxiety in Adolescent and Adult Women: A Systematic Review for the Women's Preventive Services Initiative. Ann Intern Med 2020;173:29–41. [PubMed: 32510989]

[69]. Kroenke K Two Birds with One Stone: Joint Screening for Perinatal Depression and Anxiety. J Womens Health (Larchmt) 2020.

[70]. Buttner MM, Stuart S and O'Hara MW. Moving beyond the EPDS: The association between postpartum depressive symptoms and premenstrual dysphoric disorder. Archives of Women's Mental Health 2011;14:S62.

[71]. Mendes AV, Martín-Santos R, Crippa JAS and Loureiro SR. A case-control study of school age children of depressed mothers. Archives of Women's Mental Health 2011;14:S158.

[72]. Sheeber LB, Feil EG, Seeley JR, Leve C, Gau JM, Davis B, et al. Mom-net: Evaluation of an internet-facilitated cognitive behavioral intervention for low-income depressed mothers. Journal of Consulting and Clinical Psychology 2017;85:355–366. [PubMed: 28333536]

[73]. Gliklich RE, Leavy MB, Cosgrove L, Simon GE, Gaynes BN, Peterson LE, et al. Harmonized Outcome Measures for Use in Depression Patient Registries and Clinical Practice. Ann Intern Med 2020;172:803–809. [PubMed: 32422056]

[74]. Chae SY, Chae MH, Tyndall A, Ramirez MR and Winter RO. Can we effectively use the two-item PHQ-2 to screen for postpartum depression? Fam Med 2012;44:698–703. [PubMed: 23148001]

[75]. Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, et al. 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 2020;323:2290–2300. [PubMed: 32515813]

[76]. Littlewood E, Ali S, Dyson L, Keding A, Ansell P, Bailey D, et al. Identifying perinatal depression with case-finding instruments: a mixed-methods study (BaBY PaNDA - Born and Bred in Yorkshire PeriNatal Depression Diagnostic Accuracy). Health Services Research Delivery 2018;6.

[77]. Howard LM, Ryan EG, Trevillion K, Anderson F, Bick D, Bye A, et al. Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. British Journal of Psychiatry 2018;212:50–56.

[78]. Zhong Q-Y, Gelaye B, Rondon MB, Sánchez SE, Simon GE, Henderson DC, et al. Using the Patient Health Questionnaire (PHQ-9) and the Edinburgh Postnatal Depression Scale (EPDS) to assess suicidal ideation among pregnant women in Lima, Peru. Archives of Women's Mental Health 2015;18:783–792.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[79]. Shulman B, Dueck R, Ryan D, Breau G, Sadowski I and Misri S. Feasibility of a mindfulness-based cognitive therapy group intervention as an adjunctive treatment for postpartum depression and anxiety. Journal of Affective Disorders 2018;235:61–67. [PubMed: 29653295]

[80]. Seino H, Kokai M and Matsunaga H. Effectiveness of comprehensive interprofessional intervention program (CIIP) for pregnant women with psychiatric disorders during perinatal period. Archives of Women's Mental Health 2015;18:331.

[81]. Huang MY, Suthoff E, Deligiannidis K, Lasser R, Gunduz-Bruce H, Silber C, et al. 934: Phase 3, randomized, placebo-controlled trial of SAGE-217 in postpartum depression: Association between HAM-D and PHQ-9. American Journal of Obstetrics and Gynecology 2020;222:S578.

[82]. Yawn BP, Bertram S, Kurland M and Wollan P. Anxiety assessment should be added to postpartum depression screening: A research letter. Journal of Women's Health 2015;24:689–690.

[83]. Accortt EE and Wong MS. It Is Time for Routine Screening for Perinatal Mood and Anxiety Disorders in Obstetrics and Gynecology Settings. Obstetrical and Gynecological Survey 2017;72:553–568. [PubMed: 28905985]

[84]. Falah-Hassani K, Shiri R and Dennis CL. The prevalence of antenatal and postnatal co-morbid anxiety and depression: a meta-analysis. Psychol Med 2017;47:2041–2053. [PubMed: 28414017]

[85]. Gregory KD, Chelmow D, Nelson HD, Van Niel MS, Conry JA, Garcia F, et al. Screening for Anxiety in Adolescent and Adult Women: A Recommendation From the Women's Preventive Services Initiative. Ann Intern Med 2020;173:48–56. [PubMed: 32510990]

[86]. Simpson W, Glazer M, Michalski N, Steiner M and Frey BN. Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. Can J Psychiatry 2014;59:434–40. [PubMed: 25161068]

[87]. Barrera AZ, Moh YS, Nichols A and Le HN. The Factor Reliability and Convergent Validity of the Patient Health Questionnaire-4 Among an International Sample of Pregnant Women. J Womens Health (Larchmt) 2020.

[88]. Owora AH, Carabin H, Reese J and Garwe T. Summary diagnostic validity of commonly used maternal major depression disorder case finding instruments in the United States: A meta-analysis. J Affect Disord 2016;205:335–343. [PubMed: 27566453]

[89]. Moraes GPdA, Lorenzo L, Pontes GAR, Montenegro MC and Cantilino A. Screening and diagnosing postpartum depression: When and how? Trends in Psychiatry and Psychotherapy 2017;39:54–61. [PubMed: 28403324]

[90]. Lyubenova A, Neupane D, Levis B, Wu Y, Sun Y, He C, et al. Depression prevalence based on the Edinburgh Postnatal Depression Scale compared to Structured Clinical Interview for DSM DIsorders classification: Systematic review and individual participant data meta-analysis. Int J Methods Psychiatr Res 2020:e1860. [PubMed: 33089942]

[91]. Grote NK, Katon WJ, Russo JE, Lohr MJ, Curran M, Galvin E, et al. A Randomized Trial of Collaborative Care for Perinatal Depression in Socioeconomically Disadvantaged Women: The Impact of Comorbid Posttraumatic Stress Disorder. J Clin Psychiatry 2016;77:1527–1537. [PubMed: 28076671]

[92]. Fisher J, Rowe H, Wynter K, Tran T, Lorgelly P, Amir LH, 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.

[93]. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: A randomized controlled trial of MumMoodBooster. Journal of Medical Internet Research 2016;18.

[94]. Weobong B, ten Asbroek AH, Soremekun S, Manu AA, Owusu-Agyei S, Prince M, et al. Association of antenatal depression with adverse consequences for the mother and newborn in rural Ghana: findings from the DON population-based cohort study. PLoS One 2014;9:e116333.

[95]. Yawn BP, Bertram S, Kurland M and Wollan PC. Repeated depression screening during the first postpartum year. Annals of Family Medicine 2015;13:228–234. [PubMed: 25964400]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[96]. Sidebottom A, Vacquier M, LaRusso E, Erickson D and Hardeman R. Perinatal depression screening practices in a large health system: identifying current state and assessing opportunities to provide more equitable care. Arch Womens Ment Health 2020.

[97]. Flanagan T and Avalos LA. Perinatal obstetric office depression screening and treatment. Obstetrics and Gynecology 2016;127:911–915. [PubMed: 27054937]

[98]. Santos IS, Tavares BF, Munhoz TN, Manzolli P, de Ávila GB, Jannke E, et al. Patient Health Questionnaire-9 versus Edinburgh Postnatal Depression Scale in screening for major depressive episodes: a cross-sectional population-based study. BMC Res Notes 2017;10:57. [PubMed: 28109311]

[99]. Lai BPY, Tang AKL, Lee DTS, Yip ASK and Chung TKH. Detecting postnatal depression in Chinese men: A comparison of three instruments. Psychiatry Research 2010;180:80–85. [PubMed: 20493548]

[100]. Konishi M, Tachibana Y, Tang J, Takehara K, Kubo T, Hashimoto K, et al. A comparison of self-rated and female partner-rated scales in the assessment of paternal prenatal depression. Community Mental Health Journal 2016;52:983–988. [PubMed: 26308837]

[101]. Underwood L, Waldie KE, Peterson E, D'Souza S, Verbiest M, McDaid F, et al. Paternal depression symptoms during pregnancy and after childbirth among participants in the Growing Up in New Zealand study. JAMA Psychiatry 2017;74:360–369. [PubMed: 28199455]

[102]. Pérez F, Brahm P, Riquelme S, Rivera C, Jaramillo K and Eickhorst A. Paternal post-partum depression: How has it been assessed? A literature review. Mental Health and Prevention 2017;7:28–36.

[103]. January J and Chimbari MJ. Study protocol on criterion validation of Edinburgh Postnatal Depression Scale (EPDS), Patient Health Questionnaire (PHQ-9) and Centre for Epidemiological Studies-Depression (CES-D) screening tools among rural postnatal women; a cross-sectional study. BMJ Open 2018;8:e019085.

[104]. Cena L, Palumbo G, Mirabella F, Gigantesco A, Stefana A, Trainini A, et al. Perspectives on Early Screening and Prompt Intervention to Identify and Treat Maternal Perinatal Mental Health. Protocol for a Prospective Multicenter Study in Italy. Front Psychol 2020;11:365. [PubMed: 32218756]

[105]. Stewart DE, Vigod S. Postpartum depression. N Engl J Med 2016;375:2177–86. [PubMed: 27959754]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- The PHQ-9 has good sensitivity and specificity in screening for perinatal depression
- Operating characteristics of the PHQ-9 are similar to the Edinburgh Postnatal Depression Scale
- Both scales are viable options for use in women during pregnancy and the postpartum period.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

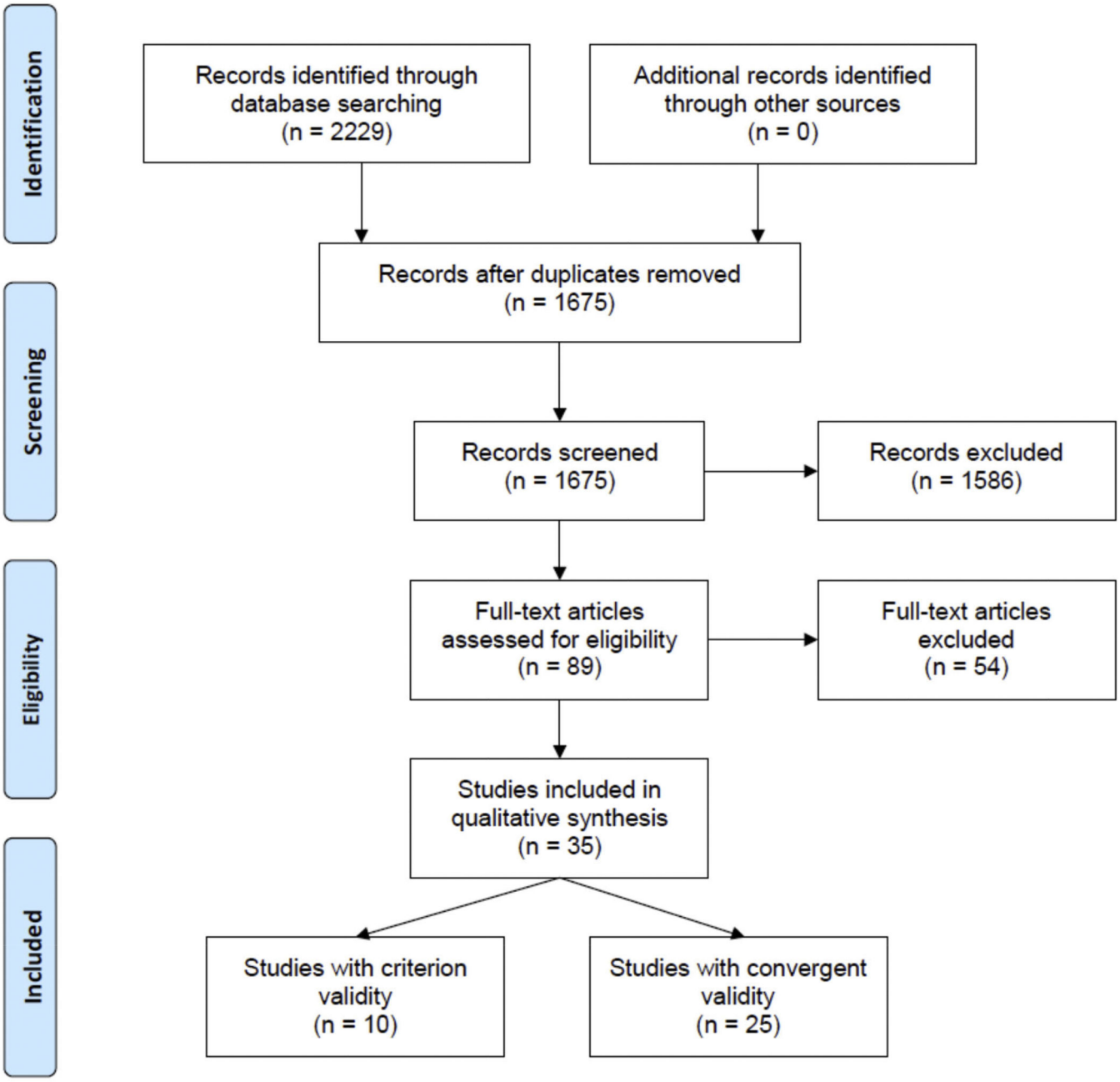

Figure 1.
PRISMA flow diagram of study selection

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

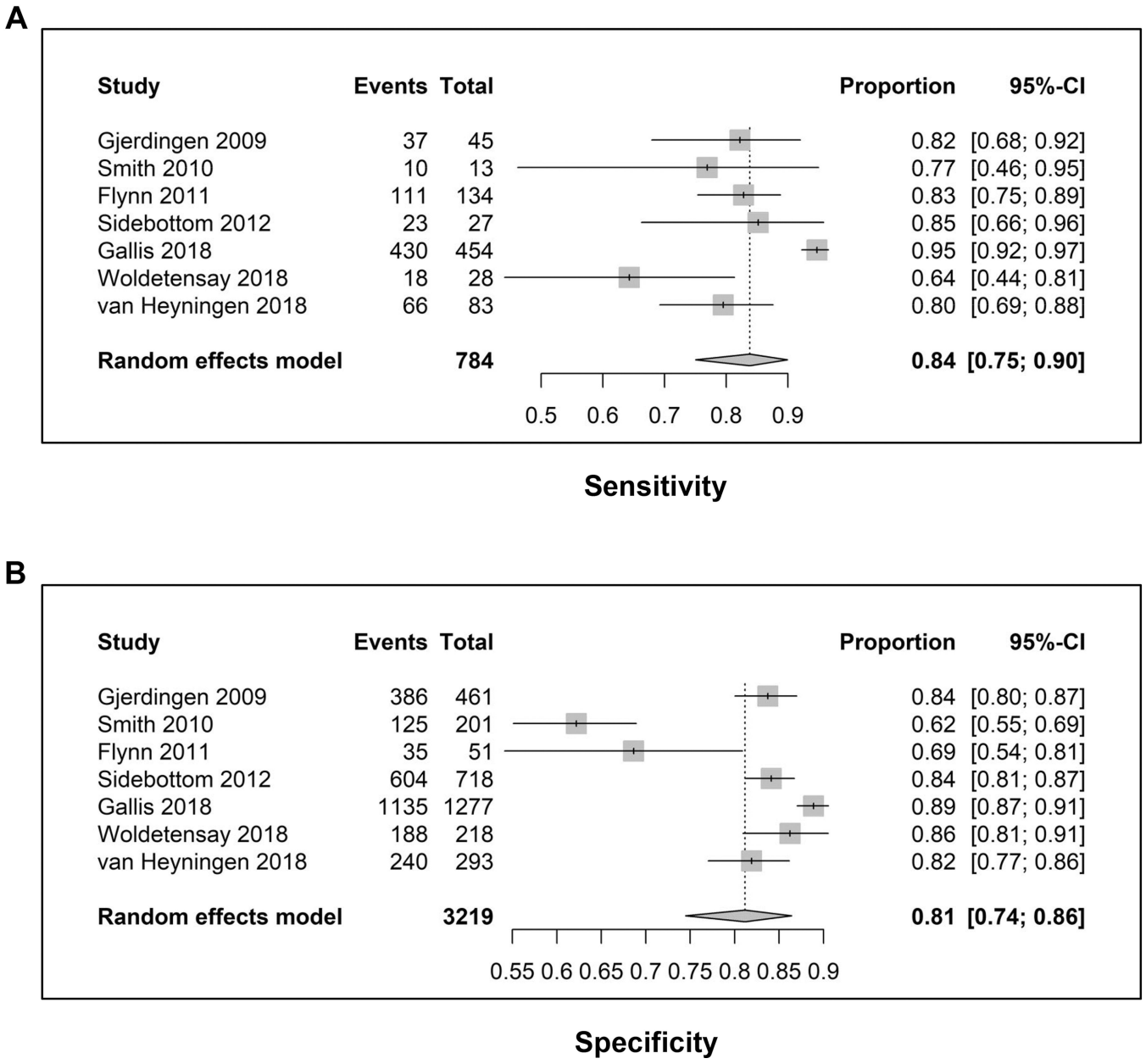

Figure 2.

Forest plots: Sensitivity (A) and specificity (B) of the 7 diagnostic accuracy studies that used the standard PHQ-9 cutpoint of a score ≥ 10. For sensitivity, total is the number of patients with major depression by the criterion standard interview and events is the number of these patients with a PHQ-9 score ≥ 10 (i.e., true positives). For specificity, total is the number of patients without major depression by the criterion standard interview and events is the number of these patients with a PHQ-9 score < 10 (i.e., true negatives).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Studies Examining Criterion or Convergent Validity of PHQ-9 in Perinatal Depression Screening

Author Year	N	Pregnant	Postpartum	Age *mean*	Country	Setting*	PHQ ≥10 %	EPDS	Validity Results†
Barthel 2015	1024	X		28.7	Ghana & Cote I'Voire	OB Hospital	30.6		Correlation of PHQ-9 with WHO-DAS disability = 0.41. Confirmatory factorial validity
Beck 2012	80		X	24.7	USA	PP Clinic	13.8		Correlation of PHQ-9 and PDSS = 0.65 Moderate concordance (no, mild, moderate to severe depression): weighted kappa= 0.40
Brodey 2016	879	X	X	27.6	USA	OB Clinics		X	See Table S1
Buttner 2013	478		X	29.6	USA	Phone survey	100		Of 478 who had PHQ-9 ≥ 10 and a SCID, PPV = 29.1%
Davis 2013	1392		X	28.5	USA	Phone survey	54.1		Of 1011 who had PHQ-9 ≥10 and a SCID, PPV of PHQ-9 ≥10 = 54%. AUC = 0.826
Di Venanzio 2017	225	X	X	33.9	Italy	Psych and OB	31.1		Of 70 who had PHQ-9 ≥5 and a psychiatric interview PPV of PHQ-9 ≥5 = 56%
Flynn 2011	185	X	X	28.2	USA	Psychiatry		X	See Table S1
Gallis 2018	1731	X		26.7	Pakistan	Community	33		See Table S1
Gawlik 2013	273	X		32.8	Germany	OB Clinic	9	X	Of 5 patients with minor or major depression by SCID, 4 and 2 exceeded EPDS ≥ 12 and PHQ-9 ≥ 10 cutoffs. Of 266 without depression, 234 and 246 were below cutoffs
Gelaye 2017	3342	X		28.2	Peru	Cohort study		X	Correlation of PHQ-9 and EPDS = 0.51
Gjerdingen 2009	506		X	29.1	USA	Pediatrics			See Table S1
Green 2018	193	X		30.6	Kenya	Community		X	See Table S1
Hanusa 2008	135		X	29.5	USA	Phone survey	17	X	Correlation of PHQ-9 & EPDS = 0.75. By DIS, AUC of PHQ-9 & EPDS = 0.80 & 0.88.
Harrington 2018	299	X		26*	Malawi	HIV		X	Good concordance (no, mild, moderate, severe depression) of PHQ-9 and EPDS: weighted kappa= 0.53
Joshi 2020	100	X		23.5	India	AN Clinic	15	X	Excellent concordance between PHQ-9 and EPDS. Kappa = 0.76
Kadir 2009	293		X	31.5	Malaysia	Hospital		X	Correlation of PHQ-9 & EPDS = 0.36 Depression prevalence using EPDS ≥ 12 and PHQ-9 ≥5 was 22.5% and 34.8%
Kulathilaka 2016	255	X		29.6	Sri Lanka	Hospital	14.1		MDD prevalence by structured interview and PHQ-9 ≥ 10 similar (13.7% vs. 14.1%)
Lara 2015	210	X	X	29.5	Mexico	AN Clinic			PHQ-9 and SCID completed at 3 time points (3rd trimester, 6 wk. and 6 mo. postpartum) Depression by SCID was 9.0%, 13.8%, and 13.3%. Depressive symptoms (PHQ-9

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Year	N	Pregnant	Postpartum	Age *mean*	Country	Setting*	PHQ ≥10 %	EPDS	Validity Results†
									≥10) prevalence was 16.6%, 17.1% and 20.0%.
Loughnan 2019	120		X	32.6	Australia	Clinical trial		X	Responsiveness: PHQ-9 and EPDS showed similar between-group differences (effect sizes = 0.99 and 0.90).
Maliszewska 2017	548		X	30.2	Poland	Hospital	13.3	X	At 4 weeks, 48 (11.7%) patients had EPDS ≥ 13, and 61 (14.9%) had PHQ-9 ≥ 10, and 30 (7.3%) exceeded both thresholds. Correlation between the two scales was 0.70 at 4 weeks and 0.60 at 3 mo. (mean=0.65)
Meltzer-Brody 2014	91	X	X	28.1	USA	Psychiatry Inpatient		X	Responsiveness. PHQ-9 and EPDS showed large effect size changes in depression over time (1.32 and 1.85, respectively)
Miller 2012	541	X	X		USA	Family med	9		PPV = 45% (13/29) by clinical interview in those with PHQ-9 ≥ 10
Mochache 2018	255	X		20–29	Kenya	AN Clinic		X	In 153 patients with an EPDS ≥ 10 PPV of PHQ-9 ≥ 5 = 71.9%
Nieminen 2016	56		X	34.6	Sweden	Clinical trial			Responsiveness: Beck (BDI) and PHQ-9 showed similar effect size changes (0.29 and 0.20) in depressive symptoms over time
Orta 2015	1321	X		33.3	USA	Cohort study	13.7		PHQ-9 ≥ 10 = 13.7%, DASS ≥10 = 14.2% and DASS ≥14 = 5.9%
Osok 2018	176	X		18	Kenya	AN Clinic	54.5	X	At least moderate depression was present in 55% by PHQ-9 ≥10 and 58% by EPDS ≥13.
Sanchez 2013	959		X	28.3	Peru	Hospital	7.4		At least moderate depression was present in 7.4% by PHQ-9 ≥10 and 7.6% by DASS ≥14
Sefogah 2020	350		X	20–34	Ghana	PP Clinic		X	Of those with PHQ-9 ≥ 5 (n=350), 32.6% had EPDS ≥ 10.
Sidebottom 2012	745	X		23	USA	AN Clinics	18		See Table S1
Smith 2010	218	X		28.9	USA	OB Clinics	5		See Table S1
van Heyningen 2018	376	X	X	26.8	S. Africa	AN Clinic		X	See Table S1
Weobong 2009	160		X	27.1	Ghana	Clinical Trial		X	See Table S1
Woldetensay 2018	246	X		24.3	Ethiopia	Community	18		See Table S1
Yawn 2009	481		X	25–29	USA	Family med	19	X	EPDS and PHQ-9 concordant in 399 women (83%) including 326 normal on both scales, and 73 elevated on both
Zhong 2014	1517	X		28	Peru	AN Clinic	29	X	Correlation of PHQ-9 & EPDS = 0.52. At score ≥ 10 on PHQ-9 & EPDS, 29% and 28% exceeded cutpoint. Agreement between 2 scales at this cutpoint was 74%. Weighted kappa using severity categories = 0.35

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Abbreviations: DASS = Depression Anxiety Stress Scale. DIS = Diagnostic Interview Schedule. EPDS = Edinburgh Postpartum Depression Scale. HRSD = Hamilton Rating Scale for Depression. NHW = non-Hispanic white. PDSS = Postpartum Depression Screening Scale. PPV = positive predictive value. SCID = Structured Clinical Interview for DSM.

*AN = antenatal. PP = postpartum. OB = obstetrics

†Criterion validity studies (n =10) are summarized in Table S1. The other 25 studies examine convergent validity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2.

Comparison of PHQ-9 and EPDS Psychometrics

Psychometric	Total # Studies	PHQ-9 *Median (Mean)*	EPDS *Median (Mean)*	Individual Study	PHQ-9	EPDS
Sensitivity	5	.81 (.81)	.82 (.81)			
				Flynn	.82	.87
				Van Heyningen	.79	.86
				Weobong	.94	.78
				Green	.70	.70
				Brodey	.81	.82
Specificity	5	.75 (.76)	.73 (.74)			
				Flynn	.69	.62
				Van Heyningen	.82	.81
				Weobong	.75	.73
				Green	.74	.72
				Brodey	.79	.81
AUC	5	.86 (.85)	.88 (.86)			
				Flynn	.86	.89
				Van Heyningen	,89	,91
				Weobong	.90	.84
				Green	.79	.80
				Hanusa	.80	.88
Correlation	6	.59 (.59)				
				Flynn	.75	
				Hanusa	.75	
				Maliszewska	.65	
				Zhong	.52	
				Gelaye	.51	
				Kadir	.36	
Categorical agreement	4	*				
Simple agreement				Yawn	.83	
Kappa, unweighted				Joshi	.76	
Kappa, weighted				Harrington	.53	
Kappa, weighted				Zhong	.35	
Responsiveness (ES)†	2	*				
				Loughnan	.99	.90
				Meltzer	1.32	1.85
Moderate dep., %	2	*		Osok	.55	.58
				Malis	.15	.12
Test-retest reliability	1	*		Weobong	.75	.51

*Mean (median) not calculated for psychometrics examined in only 1–2 studies or for concordance which was measured in different ways among studies

[†]Change over time measured in effect size (ES) for which 0.2, 0.5, and 0.8 represent small, moderate and large changes.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript