



OPEN ACCESS



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

Brooke Levis,^{1,2,3} Zelalem Negeri,^{1,2} Ying Sun,¹ Andrea Benedetti,^{2,4,5} Brett D Thombs,^{1,2,5,6,7,8,9} on behalf of the DEPRESSION Screening Data (DEPRESSD) EPDS Group

For numbered affiliations see end of the article.

Correspondence to: B D Thombs, Jewish General Hospital, 4333 Cote Ste Catherine Road, Montreal, Quebec, Canada, H3T 1E4
brett.thombs@mcgill.ca
(ORCID 0000-0002-5644-8432)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2020;371:m4022
<http://dx.doi.org/10.1136/bmj.m4022>

Accepted: 10 September 2020

ABSTRACT OBJECTIVE

To evaluate the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression in pregnant and postpartum women.

DESIGN

Individual participant data meta-analysis.

DATA SOURCES

Medline, Medline In-Process and Other Non-Indexed Citations, PsycINFO, and Web of Science (from inception to 3 October 2018).

ELIGIBILITY CRITERIA FOR SELECTING STUDIES

Eligible datasets included EPDS scores and major depression classification based on validated diagnostic interviews. Bivariate random effects meta-analysis was used to estimate EPDS sensitivity and specificity compared with semi-structured, fully structured (Mini International Neuropsychiatric Interview (MINI) excluded), and MINI diagnostic interviews separately using individual participant data. One stage meta-regression was used to examine accuracy by reference standard categories and participant characteristics.

RESULTS

Individual participant data were obtained from 58 of 83 eligible studies (70%; 15 557 of 22 788 eligible

participants (68%), 2069 with major depression). Combined sensitivity and specificity was maximised at a cut-off value of 11 or higher across reference standards. Among studies with a semi-structured interview (36 studies, 9066 participants, 1330 with major depression), sensitivity and specificity were 0.85 (95% confidence interval 0.79 to 0.90) and 0.84 (0.79 to 0.88) for a cut-off value of 10 or higher, 0.81 (0.75 to 0.87) and 0.88 (0.85 to 0.91) for a cut-off value of 11 or higher, and 0.66 (0.58 to 0.74) and 0.95 (0.92 to 0.96) for a cut-off value of 13 or higher, respectively. Accuracy was similar across reference standards and subgroups, including for pregnant and postpartum women.

CONCLUSIONS

An EPDS cut-off value of 11 or higher maximised combined sensitivity and specificity; a cut-off value of 13 or higher was less sensitive but more specific. To identify pregnant and postpartum women with higher symptom levels, a cut-off of 13 or higher could be used. Lower cut-off values could be used if the intention is to avoid false negatives and identify most patients who meet diagnostic criteria.

REGISTRATION

PROSPERO (CRD42015024785).

Introduction

Depression is common in pregnant and postpartum women and is associated with adverse outcomes for the mother, developing child, mother-infant relationship, and intimate partner relationship.^{1 2} Depression screening could potentially improve detection and management of perinatal depression. Depression screening involves the use of self-report depression symptom questionnaires to identify women above a preidentified cut-off value for further evaluation to determine whether depression is present.^{3 4} In the United Kingdom, the National Institute for Health and Care Excellence guidelines⁵ suggest that healthcare providers consider asking pregnant or postpartum women the two Whooley questions,⁶ and administering the Edinburgh Postnatal Depression Scale (EPDS) or Patient Health Questionnaire-9 screening questionnaires as part of a full assessment if depression is suspected. The guidelines do not recommend administering a screening tool to all women. The UK National Screening Committee⁷ and Canadian Task Force on Preventive Health Care⁸ recommend against screening

WHAT IS ALREADY KNOWN ON THIS TOPIC

The Edinburgh Postnatal Depression Scale (EPDS) is the most commonly used depression screening tool in perinatal care, with cut-off values of 10 or higher and 13 or higher typically used to identify women who might be depressed. A previous meta-analysis of the screening accuracy of the EPDS, which was conducted more than 13 years ago, found that a cut-off value of 12 or higher maximised combined sensitivity and specificity in postpartum women (21 studies).

The previous meta-analysis did not pool results for pregnant women because too few studies were found, and no subgroup analyses were conducted among postpartum women because primary studies did not report the necessary data.

WHAT THIS STUDY ADDS

An EPDS cut-off value of 11 or higher maximised combined sensitivity and specificity (81% and 88%, respectively).

For commonly used cut-off values of 10 or higher and 13 or higher, sensitivity and specificity were 85% and 84%, and 66% and 95%, respectively; results did not differ across subgroups, including pregnant versus postpartum status.

An online knowledge translation tool is available to estimate the expected number of positive screens and true and false screening outcomes based on study results (depressionscreening100.com/epds).

owing to concerns about false positives, possible harms, and the lack of evidence from well conducted trials that screening improves mental health outcomes. However, the United States Preventive Services Task Force (USPSTF)⁹ and Australian national guidelines¹⁰ recommend depression screening in pregnant and postpartum women, although the USPSTF notes that “screening should be implemented with adequate systems in place to ensure accurate diagnosis, effective treatment, and appropriate follow-up.” Depression screening is sometimes promoted in low and middle income countries, but it is not known whether it would improve mental health in those settings.²

The 10 item EPDS is the most commonly used depression screening tool in perinatal care; cut-off values of 10 or higher and 13 or higher are most often used to identify women who might have depression.¹¹⁻¹⁵ The USPSTF recommends screening pregnant and postpartum women with the EPDS, but does not specify a cut-off value.⁹ The systematic review conducted to support the USPSTF guideline reported the range of accuracy estimates for EPDS cut-off values of 10 or higher (1 studies) and 13 or higher (17 studies) in 23 primary studies, but did not include a meta-analysis.^{14 15} An existing meta-analysis that has examined EPDS screening accuracy searched databases through February 2007 and found that combined sensitivity and specificity to detect major depression in postpartum women was greater for a cut-off value of 12 or higher (sensitivity 0.86, specificity 0.87, 15 studies) than for a cut-off value of 10 or higher (sensitivity 0.92, specificity 0.77, 14 studies) or 13 or higher (sensitivity 0.79, specificity 0.89, 18 studies) among a total of 18 studies.¹³ The results were not pooled for pregnant women because there were too few studies, and no subgroup analyses were conducted among postpartum women because primary studies did not report the necessary data. Estimates were not done separately for different types of reference standards, although important differences exist in design and structure, and in the likelihood of major depression classification between different diagnostic interviews.¹⁶⁻¹⁸ Therefore, the optimal cut-off value for screening remains unknown, and whether different cut-off values are needed for women with different characteristics needs to be determined.

Whereas conventional meta-analyses synthesise aggregate results from study reports, individual participant data meta-analysis (IPDMA) involves the synthesis of participant level data from primary studies.¹⁹ The advantages of an IPDMA of the EPDS are the ability to include data from studies that collected EPDS and reference standard outcomes but did not publish accuracy results; the results for all cut-off values from all included studies can be taken into account rather than just published cut-off results; subgroup analyses can be conducted, which were not done in primary studies; and accuracy results can be reported separately for different reference standards. Our objectives were to use IPDMA to evaluate EPDS screening accuracy among studies that used different

types of reference standards separately, with semi-structured interviews prioritised; and to investigate whether EPDS screening accuracy differs based on pregnant versus postpartum status, age, and country human development index.

Methods

This IPDMA was registered in PROSPERO (CRD42015024785), a protocol was published,²⁰ and the results were reported following PRISMA-DTA (preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies)²¹ and PRISMA-IPD (preferred reporting items for systematic review and meta-analyses of individual participant data)²² guidelines. We followed similar methods to those used in our previously published Patient Health Questionnaire-9 diagnostic accuracy IPDMA.²³ Individual prediction models described in the protocol will be developed in future database versions. Deviations from the protocol include searching from database inception rather than from year 2000, including only one assessment time point for each woman given the small number of studies with multiple time points, and reporting results for cut-off values of 7-15 rather than 9-15.

Study eligibility

Datasets from studies that met the following criteria were deemed eligible: they administered the EPDS; diagnostic classification for current major depressive disorder or major depressive episode used Diagnostic and Statistical Manual of Mental Disorders (DSM)^{24 26} or international classification of diseases (ICD)²⁷ criteria based on a validated semi-structured or fully structured interview; the EPDS and diagnostic interview were conducted no more than two weeks apart; participants were adult women aged at least 18 years who completed assessments during pregnancy or within 12 months of giving birth; and participants were not recruited because they were receiving psychiatric assessment or care, or because they were identified as having possible depression because screening seeks to identify women with otherwise unrecognised major depression.²⁸ Studies in which some participants did not meet eligibility criteria were included in the IPDMA if primary data allowed for the selection of eligible participants.

Database searches and study selection

A medical librarian designed a peer reviewed²⁹ search strategy (eMethods1 in supplementary material) and searched Medline, Medline In-Process and Other Non-Indexed Citations, and PsycINFO through OvidSP, and Web of Science through ISI Web of Knowledge from inception to 3 October 2018. Additionally, investigators examined citations from relevant reviews and requested information about unpublished studies from authors who contributed studies. Citations identified by the search were uploaded into RefWorks (RefWorks-COS, Bethesda, MD, USA). Duplicates were removed and unique citations were

uploaded into DistillerSR (Evidence Partners, Ottawa, Canada).

Two reviewers independently reviewed titles and abstracts. For publications deemed potentially eligible by either reviewer, a full text review was done by two reviewers, also independently. Any conflicts were resolved by consensus and a third reviewer was consulted if necessary.

Data contribution, extraction, and synthesis

We invited investigators with eligible datasets to contribute deidentified versions of their datasets. We attempted to contact corresponding authors of eligible primary studies by email up to three times, as necessary. When authors did not respond to our emails, we tried to contact them by phone and emailed their coauthors.

Two investigators independently extracted information on the diagnostic interview administered and the country of study from the published reports. We used the United Nation's human development index, based on year of study publication, which reflects life expectancy, education, and income,³⁰ to categorise countries as very high, high, or low-medium development. Participant level data included in the synthesised dataset included age, pregnant or postpartum status, EPDS scores, and major depression classification status. We used major depressive disorder or major depressive episode based on the DSM or ICD criteria to classify major depression; if both were reported, we used major depressive episode because screening attempts to detect episodes of depression. Additional assessment would be needed to determine if episodes are related to major depressive disorder or another psychiatric disorder (bipolar disorder, persistent depressive disorder). We also prioritised DSM over ICD. We used statistical weights to reflect sampling procedures if provided in the datasets; for instance, when primary studies administered a diagnostic interview to all participants with positive screening results but only a random sample of those with negative results. Some studies used sampling procedures that merited weights but did not use weights. For those studies, we used inverse selection probabilities to generate appropriate weights.

We verified that participant characteristics and accuracy results from individual datasets matched those that had been published. If any discrepancies were found, we worked with the primary study investigators to understand and resolve differences. All study level and individual level participant data were transformed into a standardised format and combined in a single synthesised dataset. For nine studies that collected data at multiple time points (four with two time points, four with three time points, and one with four time points), we selected the time point with the most participants. If the number of participants was maximised at multiple time points, we selected the one with the most women who had major depression.

Risk of bias assessment

We used the Quality Assessment of Diagnostic Accuracy Studies-2 tool (QUADAS-2; eMethods²)³¹ to assess risk of bias of included studies. Two investigators independently performed this assessment and any differences were resolved through consensus or by involving a third investigator if necessary. Values used in the risk of bias assessment were coded at both study and participant levels because some values might have differed among participants from the same study (eg, time interval between index test and reference standard).

Statistical analyses

We estimated sensitivity and specificity for three reference standard categories separately across cut-off values of 7-15 for all women. Reference standard categories included semi-structured interviews (Structured Clinical Interview for DSM Disorders (SCID),³² Clinical Interview Schedule,³³ Diagnostic Interview for Genetic Studies³⁴), fully structured interviews, excluding the Mini International Neuropsychiatric Interview (MINI³⁵;³⁶ Composite International Diagnostic Interview (CIDI),³⁷ Clinical Interview Schedule-Revised³⁸), and the MINI. We analysed studies that used different types of reference standards separately because we previously found that, controlling for depressive symptom levels, the MINI might classify depression more than other diagnostic interviews, and the CIDI might classify more participants with low level symptoms as having depression but fewer with high level symptoms.¹⁶⁻¹⁸ These findings are consistent with the design of the different types of diagnostic interviews. Semi-structured interviews are designed to be administered by an experienced diagnostician who can incorporate probes and queries, and use clinical judgment. Fully structured interviews are entirely scripted so that they can be administered by a trained lay interviewer and reduce required resources. By design, fully structured interviews are intended to increase standardisation, but this could be at the cost of reduced validity.³⁹⁻⁴² The MINI is a brief version of a fully structured interview that was designed for rapid administration and tends to be overinclusive.³⁶

We fit bivariate random effects models using Gauss-Hermite quadrature for each reference standard category, for cut-off values of 7-15 separately.⁴³ This is a two stage meta-analytic approach that models sensitivity and specificity simultaneously and accounts for the correlation between them, and for precision of estimates within studies (that is, the clustering). This model provided estimates of pooled sensitivity and specificity for each analysis. We found four studies for the fully structured subgroup, one of which included only one participant with major depression. For this subgroup, we modified the bivariate model by setting the correlation between random effects to zero, and excluded the major depression case from the study that had only one major depression case. Therefore,

we used three studies to evaluate sensitivity and four studies to measure specificity.

We constructed empirical receiver operating characteristic curves based on pooled sensitivity and specificity estimates, and calculated area under the curves for each reference standard category. Additionally, we conducted one stage meta-regressions with interactions between reference standard category (reference category: semi-structured) and accuracy coefficients (logit(sensitivity) and logit(1-specificity)). We generated nomograms to present positive and negative predictive values for the optimal cut-off value (maximising Youden's $J = \text{sensitivity} + \text{specificity} - 1$) and the commonly used cut-off values of 10 or higher and 13 or higher for assumed major depression prevalence of 5-25%.

We evaluated heterogeneity for each reference standard category by generating sensitivity and specificity forest plots for each study for the optimal cut-off value and for cut-off values of 10 or higher and 13 or higher. We quantified heterogeneity by reporting estimated variances of the random effects for sensitivity and specificity (τ^2) and by estimating I^2 , which is the ratio of the estimated standard deviation of the pooled sensitivity (or specificity) from the random effects model to that from the corresponding fixed effects model.⁴⁴

Within semi-structured and MINI reference standard categories separately, we fit one stage meta-regressions where we interacted all participant characteristics (age (measured continuously), pregnant ν postpartum status (reference category=pregnant), and country human development index (reference category=very high)) with logit(sensitivity) and logit(1-specificity). Too few studies existed that used other fully structured interviews to enable us to perform meta-regressions. We conducted post hoc analyses in which we fit additional one stage meta-regressions for year of study publication. No participants had missing data for any covariates in the meta-regressions. We assessed characteristics one at a time because models attempting to fit all participant characteristics simultaneously did not converge.

When characteristics were significantly associated with sensitivity or specificity for all or most cut-off values in the meta-regressions, we fit bivariate random effects models for cut-off values of 7-15 for each subgroup. Age was fit continuously in the meta-regression but was dichotomised ($<25 \nu \geq 25$ years⁴⁵) to estimate accuracy by subgroups. For analyses in the age less than 25 subgroup, we excluded four semi-structured studies and four MINI studies that did not have any participants with major depression because the bivariate random effects model could not be applied. Therefore, 21 participants (1%) younger than 25 were excluded from semi-structured studies, and 77 (9%) from MINI studies.

In sensitivity analyses, we conducted additional meta-regressions based on QUADAS-2 scores in semi-structured and MINI reference standard categories separately. We interacted QUADAS-2 domain scores

with logit(sensitivity) and logit(1-specificity) for all domain scores with at least 100 participants with major depression and 100 without major depression among those categorised as having low risk of bias and among those with high or unclear risk of bias. We again assessed items one at a time. We performed additional sensitivity analyses for EPDS cut-off values of 10-13 by combining IPDMA accuracy results with results from studies that did not contribute individual participant data but published eligible accuracy results.

All analyses were run in R (R version R 3.4.1, R Studio version 1.0.143) by using the glmer function within the lme4 package.

Patient and public involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community. However, an online knowledge translation tool, intended for clinicians (the end users of the EPDS screening tool), is available at depressionscreening100.com/epds. The tool allows clinicians to estimate the expected number of positive screens and true and false screening outcomes based on study results.

Results

Search results and dataset inclusion

We identified 4434 unique titles and abstracts from the database search. Of these, 4056 were excluded after title and abstract review and 257 after full text review (eTable1), resulting in 121 eligible articles from 81 unique participant samples. Of these, 56 (69%) contributed datasets (fig 1). Authors of included studies contributed data from two other studies that the search did not retrieve for a total of 58 datasets (15 557 participants, 2069 with major depression). eTable2 shows characteristics of primary studies that contributed data and eligible studies that did not provide datasets. Of 22 788 participants in 83 eligible published studies, 15 557 (68%) were included. Eligible studies that did and did not contribute data were generally similar in terms of sample size, proportion of participants with major depression (excluding non-contributing studies where number with major depression was not reported), and country human development index. The proportion of studies with pregnant women only was also similar for contributing and non-contributing studies. Among both contributing and non-contributing studies, most studies used semi-structured interviews as the reference standard, followed by the MINI, and other fully structured interviews.

Of 58 included studies, 25 included pregnant women, 30 postpartum women, and three both pregnant and postpartum women. Thirty six studies used semi-structured reference standards, including 34 that used the SCID; four used fully structured

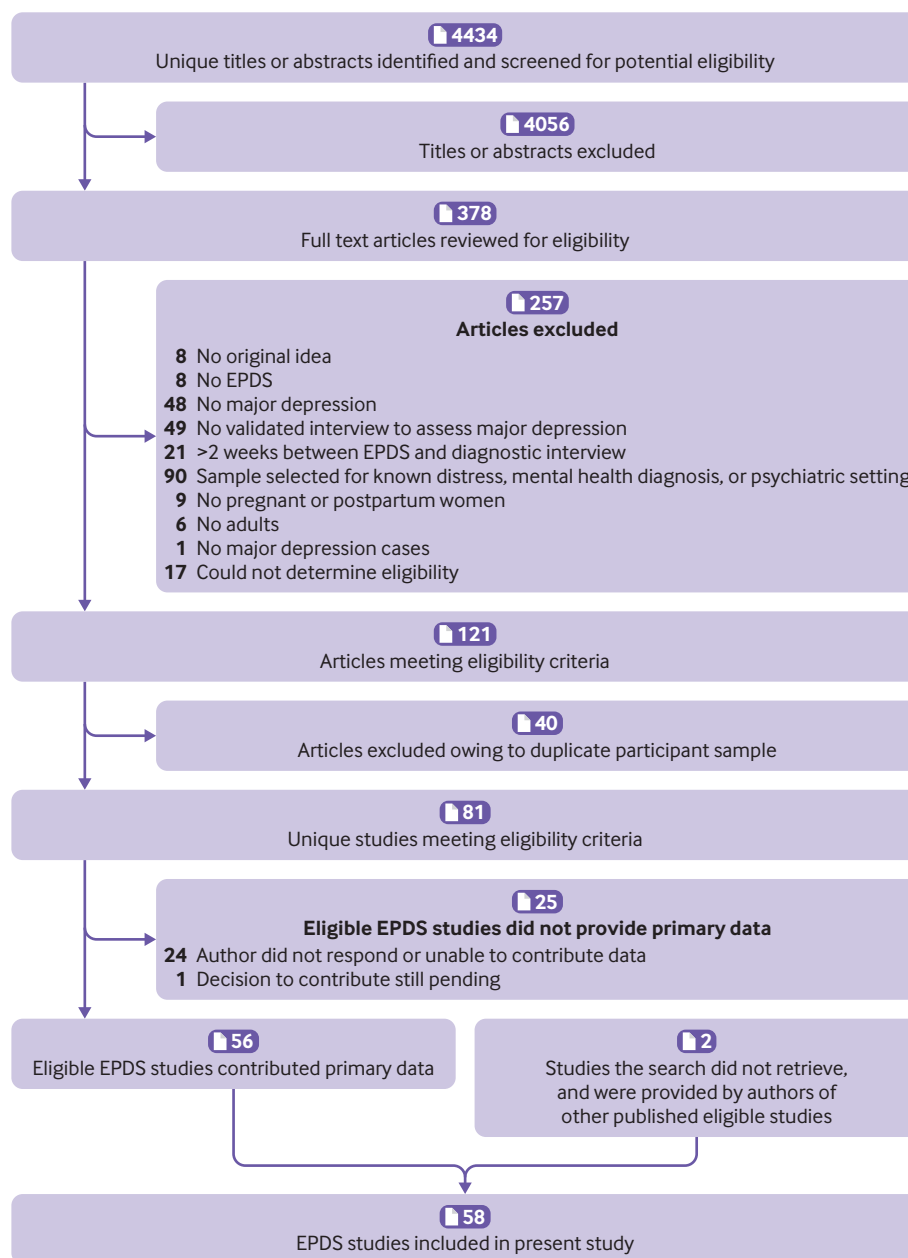


Fig 1 | Flow diagram of study selection process. EPDS=Edinburgh Postnatal Depression Scale

reference standards (MINI excluded), including three that used the CIDI; and 18 used the MINI (table 1).

EPDS sensitivity and specificity by reference standard category

Table 2 shows sensitivity and specificity estimates for cut-off values of 7-15 by reference standard category. Combined sensitivity and specificity was maximised at a cut-off value of 11 or higher for semi-structured interviews (Youden's $J=0.70$), fully structured interviews (Youden's $J=0.73$), and the MINI (Youden's $J=0.66$). For semi-structured interviews, sensitivity and specificity were 0.85 (95% confidence interval 0.79 to 0.90) and 0.84 (0.79 to 0.88) for a cut-off value of 10 or higher, 0.81 (0.75 to 0.87) and 0.88 (0.85 to 0.91) for a cut-off value of 11 or higher, and

0.66 (0.58 to 0.74) and 0.95 (0.92 to 0.96) for a cut-off value of 13 or higher. eFigure1 shows receiver operating characteristic curves and area under the curve values. No significant differences in accuracy by reference standard category were found that held across all cut-off values (eTable3). Results did not change substantively in sensitivity analyses that included published results from eight of the 26 studies that did not contribute individual participant data but published eligible accuracy results (eTable4). The other 18 eligible datasets that did not contribute individual participant data did not publish eligible diagnostic accuracy results (eTable2b).

Nomograms of positive and negative predictive values by reference standard category are shown in figure 2 (cut-off values of ≥ 11 and ≥ 13) and eFigure2

Table 1 | Participant data by diagnostic interview

| Diagnostic interview | No of studies | No of participants | No of participants with major depression (%) |
|-------------------------|---------------|--------------------|--|
| Semi-structured | | | |
| SCID | 34 | 8811 | 1292 (15) |
| CIS | 1 | 190 | 34 (18) |
| DIGS | 1 | 65 | 4 (6) |
| Fully structured | | | |
| CIDI | 3 | 2963 | 196 (7) |
| CIS-R | 1 | 226 | 32 (14) |
| MINI | 18 | 3302 | 511 (15) |
| Total | 58 | 15 557 | 2069 (13) |

CIDI=Composite International Diagnostic Interview; CIS=Clinical Interview Schedule; CIS-R=Clinical Interview Schedule-Revised; DIGS=Diagnostic Interview for Genetic Studies; DSM=Diagnostic and Statistical Manual of Mental Disorders; MINI=Mini International Neuropsychiatric Interview; SCID=Structured Clinical Interview for DSM Disorders.

(cut-off value of ≥ 10). For major depression prevalence values of 5-25% and a cut-off value of 11 or higher compared with semi-structured interviews, positive predictive values ranged from 26% to 69% and negative predictive values ranged from 93% to 99%. Ranges were similar for other reference standard types.

Heterogeneity analyses suggested moderate heterogeneity across studies. eFigure3 shows forest plots of sensitivity and specificity, and eTable5 shows τ^2 and R values.

EPDS accuracy among subgroups

Older age (measured continuously) was associated with higher specificity for both the semi-structured and MINI reference standard categories for cut-off values of 9-15 (eTable3). However, based on bivariate random effect models among participants younger than 25 and among those aged 25 or older, specificity estimates were similar across age groups (median difference across cut-off values of 7-15: 0.02 for semi-structured studies, 0.03 for MINI studies; eTable6). No other study or participant characteristics were consistently associated with differences in sensitivity or specificity estimates across both reference standard categories.

Risk of bias sensitivity analyses

eTable7 shows QUADAS-2 ratings for included studies. No QUADAS-2 domain items were consistently associated with differences in sensitivity or specificity estimates across semi-structured and MINI reference standard categories (eTable3).

Discussion

Principal findings

Our main finding was that combined sensitivity and specificity was maximised at a cut-off value of 11 or higher across reference standards. For semi-structured interviews, which are designed to closely replicate clinical diagnoses by mental health professionals, sensitivity and specificity were 81% and 88% for a cut-off value of 11 or higher. At cut-off values of 10 or higher and 13 or higher, which are commonly used for depression screening,¹³ sensitivity and specificity were 85% and 84%, and 66% and 95%, respectively. Accuracy was similar across reference standards, similar among pregnant and postpartum women, and similar based on other study and participant characteristics.

Comparison with other studies

The cut-off value of 11 or higher that maximised combined sensitivity and specificity in the present study is lower than both the most commonly used cut-off value of 13 or higher¹³ and the cut-off value of 12 or higher that maximised combined sensitivity and specificity in a previous EPDS accuracy meta-analysis.¹³ Based on studies with a semi-structured reference standard, across cut-off values of 10-13, sensitivity estimates in the present IPDMA were 6-13% lower than those in the previous meta-analysis, whereas specificity estimates were 4-7% higher. Differences in results between the current IPDMA and the previous meta-analysis might have occurred

Table 2 | Comparison of sensitivity and specificity estimates for each reference standard category

| Cut-off value | Semi-structured reference standard* | | Fully structured reference standard† | | MINI reference standard‡ | |
|---------------|-------------------------------------|----------------------|--------------------------------------|----------------------|--------------------------|----------------------|
| | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
| 7 | 0.95 (0.91 to 0.97) | 0.65 (0.58 to 0.71) | 0.95 (0.71 to 0.99) | 0.57 (0.36 to 0.76) | 0.95 (0.89 to 0.98) | 0.60 (0.52 to 0.67) |
| 8 | 0.92 (0.87 to 0.95) | 0.72 (0.66 to 0.78) | 0.95 (0.70 to 0.99) | 0.62 (0.41 to 0.80) | 0.91 (0.85 to 0.95) | 0.67 (0.60 to 0.74) |
| 9 | 0.89 (0.83 to 0.93) | 0.78 (0.73 to 0.83) | 0.95 (0.64 to 1.00) | 0.71 (0.50 to 0.85) | 0.88 (0.80 to 0.93) | 0.74 (0.66 to 0.80) |
| 10 | 0.85 (0.79 to 0.90) | 0.84 (0.79 to 0.88) | 0.93 (0.64 to 0.99) | 0.78 (0.57 to 0.90) | 0.84 (0.74 to 0.91) | 0.79 (0.73 to 0.84) |
| 11 | 0.81 (0.75 to 0.87) | 0.88 (0.85 to 0.91) | 0.90 (0.58 to 0.98) | 0.83 (0.62 to 0.94) | 0.82 (0.71 to 0.89) | 0.84 (0.79 to 0.89) |
| 12 | 0.75 (0.67 to 0.81) | 0.92 (0.89 to 0.94) | 0.81 (0.56 to 0.94) | 0.86 (0.70 to 0.94) | 0.74 (0.60 to 0.85) | 0.89 (0.83 to 0.92) |
| 13 | 0.66 (0.58 to 0.74) | 0.95 (0.92 to 0.96) | 0.79 (0.50 to 0.94) | 0.90 (0.75 to 0.96) | 0.69 (0.54 to 0.81) | 0.91 (0.87 to 0.94) |
| 14 | 0.58 (0.50 to 0.66) | 0.96 (0.95 to 0.98) | 0.77 (0.43 to 0.94) | 0.93 (0.82 to 0.98) | 0.60 (0.45 to 0.73) | 0.94 (0.91 to 0.96) |
| 15 | 0.51 (0.44 to 0.58) | 0.97 (0.96 to 0.98) | 0.66 (0.37 to 0.87) | 0.95 (0.86 to 0.99) | 0.52 (0.39 to 0.64) | 0.95 (0.92 to 0.97) |

MINI=Mini International Neuropsychiatric Interview.

*Number of studies=36; number of participants=9066; number of participants with major depression=1330.

†Number of studies=3 for sensitivity and 4 for specificity; number of participants=3188; number of participants with major depression=227. We modified the bivariate model by setting the correlation between random effects to zero and excluded the participant with major depression from the study that had only one participant with major depression.

‡Number of studies=18; number of participants=3302; number of participants with major depression=511.

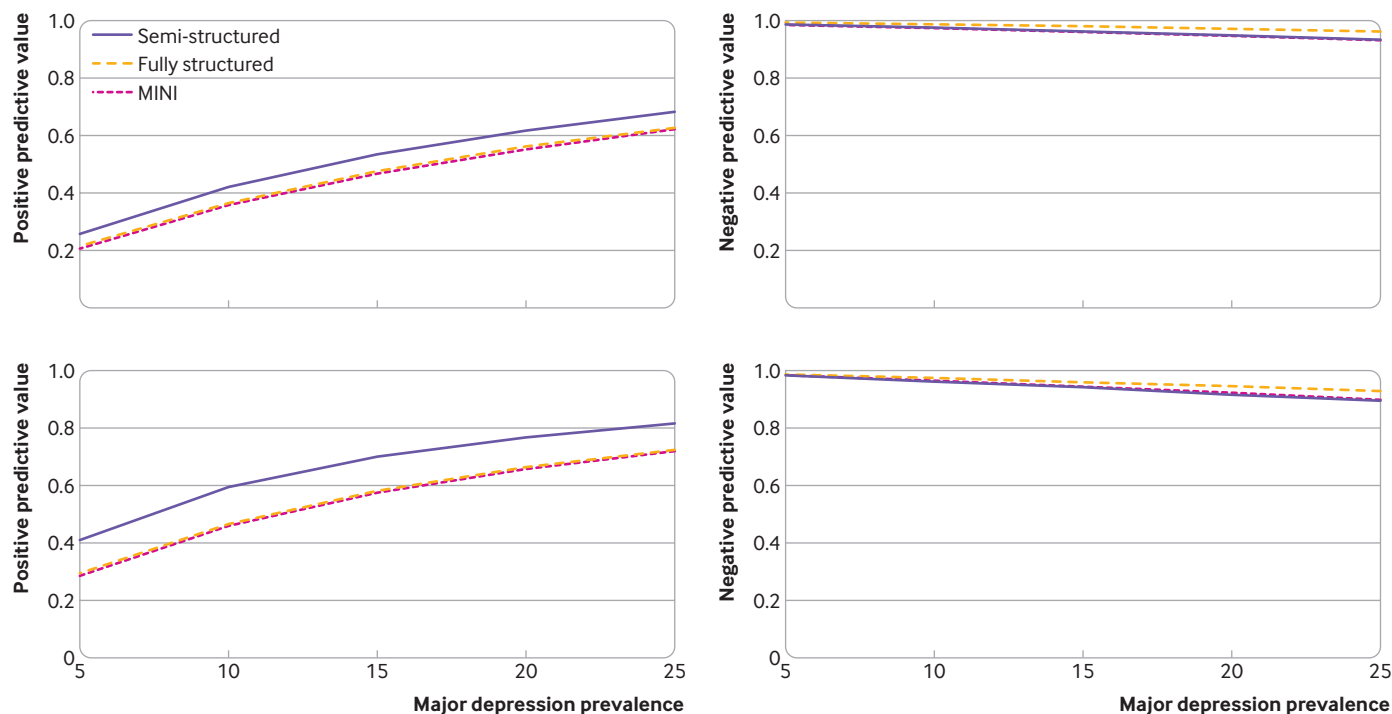


Fig 2 | Nomograms of positive and negative predictive values by reference standard category (semi-structured diagnostic interviews, fully structured diagnostic interviews, and the MINI) for major depression prevalence values of 5-25%. Upper left panel: EPDS cut-off value of 11 or higher positive predictive value; upper right panel: EPDS cut-off value of 11 or higher negative predictive value; lower left panel: EPDS cut-off value of 13 or higher positive predictive value; lower right panel: EPDS cut-off value of 13 or higher negative predictive value. EPDS=Edinburgh Postnatal Depression Scale; MINI=Mini International Neuropsychiatric Interview

because the current IPDMA consisted of 58 primary studies, including 36 with a semi-structured reference standard, versus 21 primary studies with various types of reference standards in the previous meta-analysis. Additionally, the current IPDMA incorporated data from all cut-off values for all included studies, whereas the previous meta-analysis was limited to published results and used different sets of studies to evaluate accuracy at different cut-off values.

Implications

Depression screening recommendations differ among prominent international guideline making bodies, and well conducted trials are needed to determine if screening would improve mental health or other important outcomes, such as child development and family outcomes. The present study found that an EPDS cut-off value of 11 or higher maximised combined sensitivity and specificity. Other cut-off values could be used in practice or clinical trials if either sensitivity or specificity is to be prioritised. For instance, if the intention is to only capture participants with high depressive symptom levels, a higher cut-off value might be desired. Conversely, if the intention is to avoid false negatives and capture all participants who might meet diagnostic criteria based on further evaluation, a lower cut-off value might be preferred. Clinicians considering screening for depression with the EPDS can refer to our online knowledge translation tool (depressionsscreening100.com/epds), which

estimates expected numbers of positive screens and true and false screening outcomes based on results from our IPDMA.

Strengths and limitations

This study used IPDMA to assess EPDS screening accuracy. Strengths include analysis of data from more than twice the number of studies compared with the previous conventional meta-analysis,¹³ and including results for all cut-off values from all studies. Additionally, our analysis examined the possible influence of study and participant characteristics on accuracy, and assessed accuracy separately across reference standards. A previous meta-analysis of the EPDS, which was a conventional meta-analysis and only included published results, was not able to incorporate results for all key cut-off values from all included studies because they were not consistently published. Furthermore, the previous meta-analysis was not able to conduct subgroup analyses by participant characteristics or reference standards.¹³ Among other important findings, the present study showed that the same cut-off value can be used in pregnant and postpartum women.

Limitations also need to be considered. Firstly, we did not obtain data from 25 of 83 published eligible datasets, although the results did not change when we incorporated published results from studies that did not contribute data but published eligible accuracy results. Secondly, moderate heterogeneity was found

across studies. Thirdly, we could not conduct subgroup analyses based on cultural aspects, such as country or language, or in specific pregnancy trimesters or postpartum periods because the data were insufficient. We found no significant or substantive differences based on country human development index, but few studies from low and middle income countries were included. Fourthly, while we categorised studies based on the interview administered, interviews might not always be used as intended; one third of studies were coded as unclear for interviewer qualification in our risk of bias assessment.

Conclusions

In summary, we found that combined sensitivity and specificity for the EPDS is maximised at a cut-off value of 11 or higher. Additionally, accuracy did not differ significantly based on reference standards or participant characteristics, including whether the EPDS is administered during pregnancy or in the postpartum period. Clinicians considering screening for depression with the EPDS can refer to our online tool (depressionscreening100.com/epds) to identify alternative cut-off values that maximise other parameters. Well conducted trials are needed to determine if screening with the EPDS could improve mental health outcomes and minimise harms and resource use.

AUTHOR AFFILIATIONS

¹Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

²Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

³Centre for Prognosis Research, School of Medicine, Keele University, Staffordshire, UK

⁴Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, Quebec, Canada

⁵Department of Medicine, McGill University, Montreal, Quebec, Canada

⁶Department of Psychiatry, McGill University, Montreal, Quebec, Canada

⁷Department of Psychology, McGill University, Montreal, Quebec, Canada

⁸Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada

⁹Biomedical Ethics Unit, McGill University, Montreal, Quebec, Canada

Members of the DEPRESSD EPDS Group: Chen He, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Ankur Krishnan, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Yin Wu, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Parash Mani Bhandari, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Dipika Neupane, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Mahrukh Imran, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Danielle B Rice, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Marleine Azar, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Tatiana A Sanchez, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Matthew J Chiovitti, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Nazanin Saadat, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Kira E Riehm, Lady Davis Institute for Medical Research, Jewish General

Hospital, Montréal, Québec, Canada; Jill T Boruff, Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, Québec, Canada; Lorie A Kloda, Library, Concordia University, Montréal, Québec, Canada; Pim Cuijpers, Department of Clinical, Neuro and Developmental Psychology, EMGO Institute, Vrije Universiteit Amsterdam, Netherlands; Simon Gilbody, Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK; John P A Ioannidis, Department of Medicine, Department of Epidemiology and Population Health, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, CA, USA; Dean McMillan, Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK; Scott B Patten, Departments of Community Health Sciences and Psychiatry, University of Calgary, Calgary, Canada; Ian Shrier, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; Roy C Ziegelstein, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; Liane Comeau, International Union for Health Promotion and Health Education, École de santé publique de l'Université de Montréal, Montréal, Québec, Canada; Nicholas D Mitchell, Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada; Marcello Tonelli, Department of Medicine, University of Calgary, Calgary, Alberta, Canada; Simone N Vigod, Women's College Hospital and Research Institute, University of Toronto, Toronto, Ontario, Canada; Franca Aceti, Department of Neurology and Psychiatry, Sapienza University of Rome, Rome, Italy; Rubén Alvarado, School of Public Health, Faculty of Medicine, Universidad de Chile, Santiago, Chile; Cosme Alvarado-Esquivel, Laboratorio de Investigación Biomédica, Facultad de Medicina y Nutrición, Avenida Universidad, Dgo, Mexico; Muideen O Bakare, Child and Adolescent Unit, Federal Neuropsychiatric Hospital, Enugu, Nigeria; Jacqueline Barnes, Department of Psychological Sciences, Birkbeck, University of London, UK; Amar D Bavle, Department of Psychiatry, Rajarajeswari Medical College and Hospital, Bengaluru, Karnataka, India; Cheryl Tatano Beck, University of Connecticut School of Nursing, Mansfield, CT, USA; Carola Bindt, Department of Child and Adolescent Psychiatry, University Medical Center Hamburg-Eppendorf, Germany; Philip M Boyce, Discipline of Psychiatry, Westmead Clinical School, Sydney Medical School, University of Sydney, Sydney, Australia; Adomas Bunevicius, Neuroscience Institute, Lithuanian University of Health Sciences, Kaunas, Lithuania; Humberto Correa, Medicine Faculty – Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; Tiago Castro e Couto, Federal University of Uberlândia, Brazil; Linda H Chaudron, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA; Genesis Chorwe-Sungani, Department of Mental Health, Kamuzu College of Nursing, University of Malawi, Blantyre, Malawi; Felipe Pinheiro de Figueiredo, Department of Neurosciences and Behaviour, Ribeirão Preto Medical School, Brazil; Valsamma Eapen, University of New South Wales and Ingham Institute South West Sydney LHD, Australia; Nicolas Favez, Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland; Ethel Felice, Department of Psychiatry, Mount Carmel Hospital, Attard, Malta; Gracia Fellmeth, Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Mae Sot, Thailand; Michelle Fernandes, Faculty of Medicine, Department of Paediatrics, University of Southampton, Southampton and Nuffield Department of Women's and Reproductive Health, University of Oxford, UK; Sally Field, Perinatal Mental Health Project, Alan J. Flisher Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa; Barbara Figueiredo, School of Psychology, University of Minho, Portugal; Jane R W Fisher, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; Lluïsa Garcia-Esteve, Perinatal Mental Health Unit CLINIC-BCN, Institut Clínic de Neurociències, Hospital Clínic, Barcelona, Spain; Lisa Giardinelli, Psychiatry Unit, Department of Health Sciences, University of Florence, Firenze, Italy; Eric P Green, Duke Global Health Institute, Durham, NC, USA; Nadine Helle, Department of Child and Adolescent Psychiatry, University Medical Center Hamburg-Eppendorf, Germany; Simone Honikman, Perinatal Mental Health Project, Alan J. Flisher Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa; Louise M Howard, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; Pirjo A Kettunen, Department of General Hospital Psychiatry, North Karelia Central Hospital, Joensuu, Finland; Dina Sami Khalifa, Ahfad University for Women, Omdurman, Sudan; Jane Kohlhoff, University of New South Wales, Kensington, Australia; Laima Kusminska, Private Practice, Hamburg, Germany; Zoltán Kozinszky, Department of Obstetrics and Gynaecology, Danderyd Hospital, Stockholm, Sweden; Lorenzo Lelli, Psychiatry Unit, Department of Health Sciences,

University of Florence, Firenze, Italy; Angeliki A Leonardou, First Department of Psychiatry, Women's Mental Health Clinic, Athens University Medical School, Athens, Greece; Michael Maes, Department of Psychiatry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; Carina Y Marsay, Department of Psychiatry, University of the Witwatersrand, Johannesburg, South Africa; Pablo Martínez, Escuela de Psicología, Facultad de Humanidades, Universidad de Santiago de Chile, Santiago, Chile; Valentina Meuti, Department of Neurology and Psychiatry, Sapienza University of Rome, Rome, Italy; Sandra Nakić Radoš, Department of Psychology, Catholic University of Croatia, Zagreb, Croatia; Purificación Navarro García, Perinatal Mental Health Unit CLINIC-BCN, Institut Clínic de Neurociències, Hospital Clínic, Barcelona, Spain; Daisuke Nishi, Department of Mental Health, Graduate School of Medicine, University of Tokyo, Japan; Daniel Okitundo Luwa E-Andjafono, Unité de Neuropsychologie, Département de Neurologie, Centre Neuro-psycho-pathologique, Faculté de Médecine, Université de Kinshasa, République Démocratique du Congo; Susan J Pawlby, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; Emma Robertson-Blackmore, Halifax Health, Graduate Medical Education, Daytona Beach, FL, USA; Tamsen J Rochat, MRC/Developmental Pathways to Health Research Unit, Faculty of Health Sciences, University of Witwatersrand, South Africa; Heather J Rowe, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; Deborah J Sharp, Centre for Academic Primary Care, Bristol Medical School, University of Bristol, UK; Bonnie W M Siu, Department of Psychiatry, Castle Peak Hospital, Hong Kong SAR, China; Alkistis Skalkidou, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden; Johanne Smith-Nielsen, Center for Early Intervention and Family Studies, Department of Psychology, University of Copenhagen, Denmark; Alan Stein, Department of Psychiatry, University of Oxford, Oxford, UK; Robert C Stewart, Department of Mental Health, College of Medicine, University of Malawi, Malawi; Kuan-Pin Su, An-Nan Hospital, China Medical University and Mind-Body Interface Laboratory, China Medical University Hospital, Taiwan; Inger Sundström-Poromaa, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden; Meri Tadinac, Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia; S Darius Tandon, North western University Feinberg School of Medicine, Chicago, IL, USA; Iva Tendais, School of Psychology, University of Minho, Portugal; Pavaani Thiagayson, The Institute of Mental Health, Singapore; Annamária Töröki, Department of Emergency, University of Szeged, Hungary; Anna Torres-Giménez, Perinatal Mental Health Unit CLINIC-BCN, Institut Clínic de Neurociències, Hospital Clínic, Barcelona, Spain; Thach D Tran, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; Kylee Trevillion, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; Friday P Tungchama, Department of Psychiatry, College of Health Sciences, University of Jos, Nigeria; Katherine Turner, Epilepsy Center-Child Neuropsychiatry Unit, ASST Santi Paolo Carlo, San Paolo Hospital, Milan, Italy; Mette S Væver, Centre for Early Intervention and Family Studies, Department of Psychology, University of Copenhagen, Copenhagen, Denmark; Thandi van Heyningen, Perinatal Mental Health Project, Alan J. Flisher Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa; Johann M Vega-Dienstmaier, Facultad de Medicina Alberto Hurtado, Universidad Peruana Cayetano Heredia, Lima, Perú; Karen Wynter, School of Nursing and Midwifery, Deakin University, Melbourne, Australia; and Kimberly A Yonkers, Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA.

Contributors: BL, ABe, BDT were responsible for the study conception and design; BL, ZN, YS, BDT contributed to data extraction, coding, evaluation of included studies, and data synthesis; BL, ZN, ABe, BDT contributed to data analysis and interpretation; BL, ABe, BDT drafted the manuscript. Members of the DEPRESSD EPDS Group contributed to data extraction, coding, and synthesis: CH, AK, YW, PMBh, DNe, MI, DBR, MA, TAS, MJC, NS, KER; via the design and conduct of database searches: JTB, LAK; as members of the DEPRESSD Steering Committee, including conception and oversight of collaboration: PC, SG, JPAI, DM, SBP, IS, RCZ; as a knowledge user consultant: LC, NDM, MTo, SNV; by contributing included datasets: FA, RA, CAE, MOB, JB, ADB, CTB, CB, PMBo, ABu, HC, TceC, LHC, GCS, FpF, VE, NF, EF, GF, MF, SF, BF, JRWF, LGE, LG, EPG, NH, SH, LMH, PAK, DSK, JK, LK, ZK, LL, AAL, MM, CYM, PM, VM, SNR, PNG, DNI, DOLEA, SJP, ERB, TJR, HJR, DJS, BWMS, ASK, JSN, AST, RCS, KPS, ISP, MTa, SDT, IT, PT, AT, ATG, TDT, KTr, FPT, KTU, MSV, TvH, JMVD, KW, KAY. All authors, including group authors, provided a critical review and approved the final manuscript. AB and BDT contributed equally as co-senior authors and are the guarantors; they had full access to all the data in the study and take responsibility

for the integrity of the data and the accuracy of the data analyses. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This study was funded by the Canadian Institutes of Health Research (CIHR, KRS-140994). BL and YW were supported by Fonds de recherche du Québec – Santé (FRQS) Postdoctoral Training Fellowships. ABe and BDT were supported by FRQS researcher salary awards. PMBh was supported by a studentship from the Research Institute of the McGill University Health Centre. DNe was supported by G R Caverhill Fellowship from the Faculty of Medicine, McGill University. DBR was supported by a Vanier Canada Graduate Scholarship. MA was supported by an FRQS Masters Training Award. The primary study by Alvarado et al was supported by the Ministry of Health of Chile. The primary study by Barnes et al was supported by a grant from the Health Foundation (1665/608). The primary study by Beck et al was supported by the Patrick and Catherine Weldon Donaghue Medical Research Foundation and the University of Connecticut Research Foundation. The primary study by Helle et al was supported by the Werner Otto Foundation, the Kroschke Foundation, and the Feindt Foundation. Robertas Bunevicius (1958-2016) was principal investigator of the primary study by Bunevicius et al, but passed away and was unable to participate in this project. The primary study by Figueira et al was supported by the Brazilian Ministry of Health and by the National Counsel of Technological and Scientific Development (CNPq; grant No 403433/2004-5). The primary study by Couto et al was supported by CNPq (grant No 444254/2014-5) and the Minas Gerais State Research Foundation (FAPEMIG; grant No APQ-01954-14). The primary study by Chaudron et al was supported by a grant from the National Institute of Mental Health (grant K23 MH64476). The primary study by Chorwe-Sungani et al was supported by the University of Malawi through grant QZA-0484 NORHED 2013. The primary study by de Figueiredo et al was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo. The primary study by Tissot et al was supported by the Swiss National Science Foundation (grant 32003B 125493). The primary study by Fernandes et al was supported by grants from the Child: Care Health and Development Trust and the Department of Psychiatry, University of Oxford, UK, and by the Ashok Ranganathan Bursary from Exeter College, University of Oxford. Dr Fernandes is supported by a University of Southampton National Institute for Health Research (NIHR) academic clinical fellowship in Paediatrics. The primary study by van Heyningen et al was supported by the Medical Research Council of South Africa (fund No 415865), Cordaid Netherlands (project 103/10002 G Sub 7) and the Truworths Community Foundation Trust, South Africa. Dr van Heyningen was supported by the National Research Foundation of South Africa and the Harry Crossley Foundation (VHYTHE001/1232209). The primary study by Tendais et al was supported under the project POCI/SAU-ESP/56397/2004 by the Operational Program Science and Innovation 2010 (POCI 2010) of the Community Support Board III and by the European Community Fund FEDER. The primary study by Fisher et al was supported by a grant under the Invest to Grow Scheme from the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs. The primary study by Garcia-Estevé et al was supported by grant 7/98 from the Ministerio de Trabajo y Asuntos Sociales, Women's Institute, Spain. The primary study by Green et al was supported by a grant from the Duke Global Health Institute (453-0751). The primary study by Howard et al was supported by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (grant reference No RP-PG-1210-12002 and RP-DG-1108-10012) and by the South London Clinical Research Network. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The primary study by Kettunen et al was supported with an Annual EVO Financing (special government subsidies from the Ministry of Health and Welfare, Finland) by North Karelia Central Hospital and Päijät-Häme Central Hospital. The primary study by Phillips et al was supported by a scholarship from the National Health and Medical Research Council (NHMRC). The primary study by Roomuangwong et al was supported by the Ratchadaphiseksomphot Endowment Fund 2013 of Chulalongkorn University (CU-56-457-HR). The primary study by Marsay et al was supported by the South Africa Medical Research Council in terms of the Clinician Researcher PhD Programme. The primary study by Martínez et al was supported by Iniciativa Científica Milenio, Chile, process No IS130005 and by Fondo Nacional de Desarrollo Científico y Tecnológico, Chile, process No 1130230. The primary study by Nakić Radoš et al was supported by the Croatian Ministry of Science, Education, and Sports (134-0000000-2421). The primary study by Navarro et al was supported by grant 13/00

from the Ministry of Work and Social Affairs, Institute of Women, Spain. The primary study by Usuda et al was supported by Grant-in-Aid for Young Scientists (A) from the Japan Society for the Promotion of Science (primary investigator Daisuke Nishi), and by an intramural research grant for neurological and psychiatric disorders from the National Center of Neurology and Psychiatry, Japan. The primary study by Pawlby et al was supported by a Medical Research Council UK Project Grant (number G89292999N). Dr Robertson-Blackmore was supported by a Young Investigator Award from the Brain and Behaviour Research Foundation and NIMH grant K23MH080290. The primary study by RoCHAT et al was supported by grants from the University of Oxford (HQ5035), the Tuixen Foundation (9940), the Wellcome Trust (082384/Z/07/Z and 071571), and the American Psychological Association. Dr RoCHAT receives salary support from a Wellcome Trust Intermediate Fellowship (211374/Z/18/Z). The primary study by Rowe et al was supported by the diamond Consortium, beyondblue Victorian Centre of Excellence in Depression and Related Disorders. The primary study by Comasco et al was supported by funds from the Swedish Research Council (VR: 521-2013-2339, VR: 523-2014-2342), the Swedish Council for Working Life and Social Research (FAS: 2011-0627), the Marta Lundqvist Foundation (2013, 2014), and the Swedish Society of Medicine (SLS-331991). The primary study by Smith-Nielsen et al was supported by a grant from the charitable foundation Tryg Foundation (grant No 107616). The primary study by Prenoveau et al was supported by The Wellcome Trust (grant No 071571). The primary study by Stewart et al was supported by Professor Francis Creed's Journal of Psychosomatic Research Editorship fund (BA00457) administered through University of Manchester. The primary study by Su et al was supported by grants from the Department of Health (DOH94F044 and DOH95F022) and the China Medical University and Hospital (CMU94-105, DMR-92-92, and DMR94-46). The primary study by Tandon et al was funded by the Thomas Wilson Sanitarium. The primary study by Tran et al was supported by the Myer Foundation who funded the study under its Beyond Australia scheme. Dr Tran was supported by an early career fellowship from the Australian National Health and Medical Research Council. The primary study by Vega-Dienstmaier et al was supported by Tejada Family Foundation, and Peruvian-American Endowment. The primary study by Yonkers et al was supported by a National Institute of Child Health and Human Development grant (5 R01HD045735). References for all included studies can be found at the end of the supplementary material.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: support from Canadian Institutes of Health Research for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years with the following exceptions: MTO declares that he has received a grant from Merck Canada, outside the submitted work; SNV declares that she receives royalties from UpToDate, outside the submitted work; CTB declares that she receives royalties for her Postpartum Depression Screening Scale published by Western Psychological Services; PMB declares that he receives grants and personal fees from Servier, grants from Lundbeck, and personal fees from AstraZeneca, all outside the submitted work; LMH declares that she has received personal fees from NICE Scientific Advice, outside the submitted work; ISP declares that she has served on advisory boards and acted as invited speaker at scientific meetings for MSD, Novo Nordisk, Bayer Health Care, and Lundbeck A/S; KAY declares that she receives royalties from UpToDate, outside the submitted work. All authors declare no other relationships or activities that could appear to have influenced the submitted work. No funder had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

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

Data sharing: Requests to access data should be made to the corresponding author.

The manuscript's guarantor affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: There are no plans to disseminate the results of the research to study participants or the relevant patient community. However, an online knowledge translation tool, intended for clinicians (the end users of the EPDS screening tool), is available at depressionscreening100.com/epds. The tool allows clinicians to estimate the expected number of positive screens and true and false screening outcomes based on study results.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- 1 Vesga-López O, Blanco C, Keyes K, et al. Psychiatric disorders in pregnant and postpartum women in the United States. *Arch Gen Psychiatry* 2008;65:805-15. doi:10.1001/archpsyc.65.7.805
- 2 Howard LM, Molyneaux E, Dennis CL, RoCHAT T, Stein A, Milgrom J. Non-psychotic mental disorders in the perinatal period. *Lancet* 2014;384:1775-88. doi:10.1016/S0140-6736(14)61276-9
- 3 Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? *BMJ* 2014;348:g1253. doi:10.1136/bmj.g1253
- 4 Thombs BD, Coyne JC, Cuijpers P, et al. Rethinking recommendations for screening for depression in primary care. *CMAJ* 2012;184:413-8. doi:10.1503/cmaj.111035
- 5 National Institute for Health and Care Excellence. *Antenatal and postnatal mental health: Clinical management and service guidance*. British Psychological Society, 2014.
- 6 Whooley MA, Avins AL, Miranda J, et al. Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med* 1997;12:439-45. doi:10.1046/j.1525-1497.1997.00076.x
- 7 Hill C. *An evaluation of screening for postnatal depression against NSC criteria*. National Screening Committee, 2010.
- 8 Joffres M, Jaramillo A, Dickinson J, et al. Recommendations on screening for depression in adults. *CMAJ* 2013;185:775-82. doi:10.1503/cmaj.130403
- 9 Siu ALUS Preventive Services Task Force (USPSTF). Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA* 2016;315:380-7. doi:10.1001/jama.2015.18392
- 10 Austin MP, Highet N, Expert Working Group. *Mental health care in the perinatal period: Australian clinical practice guideline*. Centre of Perinatal Excellence, 2017.
- 11 Hewitt CE, Gilbody SM, Mann R, et al. Instruments to identify postnatal depression: which methods have been the most extensively validated, in what setting and in which language? *Int J Psychiatry Clin Pract* 2010;14:72-6. doi:10.3109/1365100903198020
- 12 Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987;150:782-6. doi:10.1192/bjp.150.6.782
- 13 Hewitt CE, Gilbody SM, Brealey S, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. *Health Technol Assess* 2009;13:147-230. doi:10.3310/hta13360
- 14 O'Connor E, Rossom RC, Henninger M, et al. *Screening for depression in adults: an updated systematic evidence review for the US Preventive Services Task Force: evidence synthesis No. 128*. [AHRQ Publication No. 14-05208-EF-1] Agency for Healthcare Research and Quality, 2016.
- 15 O'Connor E, Rossom RC, Henninger M, et al. 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. doi:10.1001/jama.2015.18948
- 16 Levis B, Benedetti A, Riehm KE, et al. Probability of major depression diagnostic classification using semi-structured vs. fully structured diagnostic interviews. *Br J Psychiatry* 2018;212:377-85. doi:10.1192/bjp.2018.54
- 17 Levis B, McMillan D, Sun Y, et al. Comparison of major depression diagnostic classification probability using the SCID, CID1 and MINI diagnostic interviews among women in pregnancy or postpartum: an individual participant data meta-analysis. *Int J Methods Psychiatr Res* 2019;28:e1803. doi:10.1002/mpr.1803
- 18 Wu Y, Levis B, Sun Y, et al. Probability of major depression diagnostic classification based on the SCID, CID1, and MINI diagnostic interviews controlling for Hospital Anxiety and Depression Scale - Depression Subscale scores: an individual participant data meta-analysis

- of 73 primary studies. *J Psychosom Res* 2020;129:109892. doi:10.1016/j.jpsychores.2019.109892
- 19 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221. doi:10.1136/bmj.c221
 - 20 Thombs BD, Benedetti A, Kloda LA, et al. Diagnostic accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for detecting major depression in pregnant and postnatal women: protocol for a systematic review and individual patient data meta-analyses. *BMJ Open* 2015;5:e009742. doi:10.1136/bmjopen-2015-009742
 - 21 McInnes MD, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA statement. *JAMA* 2018;319:388-96. doi:10.1001/jama.2017.19163
 - 22 Stewart LA, Clarke M, Rovers M, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD statement. *JAMA* 2015;313:1657-65. doi:10.1001/jama.2015.3656
 - 23 Levis B, Benedetti A, Thombs BD, et al. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ* 2019;365:11476. doi:10.1136/bmj.11476
 - 24 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-III*. 3rd ed, revised. APA, 1987.
 - 25 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-IV*. 4th ed. APA, 1994.
 - 26 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-IV*. 4th ed revised. APA, 2000.
 - 27 World Health Organization. *The ICD-10 classifications of mental and behavioural disorder. Clinical descriptions and diagnostic guidelines*. WHO, 1992.
 - 28 Thombs BD, Arthurs E, El-Baalbaki G, et al. Risk of bias from inclusion of already diagnosed or treated patients in diagnostic accuracy studies of depression screening tools: A systematic review. *BMJ* 2011;343:d4825. doi:10.1136/bmj.d4825
 - 29 McGowan J, Sampson M, Salzwedel DM, et al. PRESS – Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). CADTH. https://www.cadth.ca/sites/default/files/pdf/CP0015_PRESS_Update_Report_2016.pdf 2016
 - 30 United Nations. International human development indicators. <http://hdr.undp.org/en/countries>. 2019.
 - 31 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36. doi:10.7326/0003-4819-155-8-201110180-00009
 - 32 First MB. *Structured Clinical Interview for the DSM (SCID)*. John Wiley & Sons, 1995.
 - 33 Goldberg DP, Cooper B, Eastwood MR, et al. A standardized psychiatric interview for use in community surveys. *Br J Prev Soc Med* 1970;24:18-23. doi:10.1136/jech.24.1.18
 - 34 Nurnberger J Jr, Blehar MC, Kaufmann CA, et al. Diagnostic Interview for Genetic Studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* 1994;51:849-59, discussion 863-4. doi:10.1001/archpsyc.1994.03950110009002
 - 35 Lecrubier Y, Sheehan DV, Weiller E, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry* 1997;12:224-31. doi:10.1016/S0924-9338(97)83296-8
 - 36 Sheehan DV, Lecrubier Y, Sheehan KH, et al. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *Eur Psychiatry* 1997;12:232-41. doi:10.1016/S0924-9338(97)83297-X
 - 37 Robins LN, Wing J, Wittchen HU, et al. The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry* 1988;45:1069-77. doi:10.1001/archpsyc.1988.01800360017003
 - 38 Lewis G, Pelosi AJ, Araya R, et al. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychol Med* 1992;22:465-86. doi:10.1017/S0033291700030415
 - 39 Brugha TS, Jenkins R, Taub N, et al. A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychol Med* 2001;31:1001-13. doi:10.1017/S0033291701004184
 - 40 Brugha TS, Bebbington PE, Jenkins R. A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychol Med* 1999;29:1013-20. doi:10.1017/S0033291799008880
 - 41 Nosen E, Woody SR. Diagnostic assessment in research. In: *Handbook of research methods in abnormal and clinical psychology*. Sage, 2008.
 - 42 Kurdyak PA, Gnam WH. Small signal, big noise: performance of the CIDI depression module. *Can J Psychiatry* 2005;50:851-6. doi:10.1177/070674370505001308
 - 43 Riley RD, Dodd SR, Craig JV, et al. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med* 2008;27:6111-36. doi:10.1002/sim.3441
 - 44 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58. doi:10.1002/sim.1186
 - 45 Rubertsson C, Wickberg B, Gustavsson P, et al. Depressive symptoms in early pregnancy, two months and one year postpartum-prevalence and psychosocial risk factors in a national Swedish sample. *Arch Women Ment Health* 2005;8:97-104. doi:10.1007/s00737-005-0078-8

Web appendix: Supplementary material