

BMJ Open 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

Brett D Thombs,^{1,2,3,4,5,6,7} Andrea Benedetti,^{2,3,8} Lorie A Kloda,⁹ Brooke Levis,^{2,7} Kira E Riehm,⁷ Marleine Azar,^{2,7} Pim Cuijpers,¹⁰ Simon Gilbody,¹¹ John P A Ioannidis,¹² Dean McMillan,¹¹ Scott B Patten,¹³ Ian Shrier,^{2,7} Russell J Steele,^{14,7} Roy C Ziegelstein,¹⁵ Marcello Tonelli,¹⁶ Nicholas Mitchell,¹⁷ Liane Comeau,¹⁸ Joy Schinazi,¹⁹ Simone Vigod²⁰

To cite: 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

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-009742>).

Received 18 August 2015
Accepted 24 September 2015



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Brett D Thombs;
brett.thombs@mcgill.ca

ABSTRACT

Introduction: Studies of the diagnostic accuracy of depression screening tools often used data-driven methods to select optimal cut-offs. Typically, these studies report results from a small range of cut-off points around whatever cut-off score is identified as most accurate. When published data are combined in meta-analyses, estimates of accuracy for different cut-off points may be based on data from different studies, rather than data from all studies for each cut-off point. Thus, traditional meta-analyses may exaggerate accuracy estimates. Individual patient data (IPD) meta-analyses synthesise data from all studies for each cut-off score to obtain accuracy estimates. The 10-item Edinburgh Postnatal Depression Scale (EPDS) is commonly recommended for depression screening in the perinatal period. The primary objective of this IPD meta-analysis is to determine the diagnostic accuracy of the EPDS to detect major depression among women during pregnancy and in the postpartum period across all potentially relevant cut-off scores, accounting for patient factors that may influence accuracy (age, pregnancy vs postpartum).

Methods and analysis: Data sources will include Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO, and Web of Science. Studies that include a diagnosis of major depression based on a validated structured or semistructured clinical interview administered within 2 weeks of (before or after) the administration of the EPDS will be included. Risk of bias will be assessed with the Quality Assessment of Diagnostic Accuracy Studies-2 tool. Bivariate random-effects meta-analysis will be conducted for the full range of plausible cut-off values. Analyses will evaluate data from pregnancy and the postpartum period separately, as well as combining data from all women in a single model.

Strengths and limitations of this study

- The study will use individual patient data that will allow estimate of diagnostic accuracy for all relevant cut-off scores of the Edinburgh Postnatal Depression Scale (EPDS) using data from all patients at each cut-off score. This will overcome limitations related to selective cut-off reporting in many primary study publications.
- The study will conduct analyses that exclude patients with current diagnoses of depression or who are undergoing mental health treatment at the time of study enrolment, since these patients would not be screened in clinical practice. This will overcome potential bias in primary study publications where these patients are often included.
- The study will generate predictive models that estimate the probability that a patient has depression based on EPDS scores and relevant covariates. This will facilitate more informed clinical decision-making than can be done with standard diagnostic accuracy metrics.
- A potential limitation is that the success of the study depends on the ability to obtain large amounts of individual patient data. Thus, we do not know the proportion of eligible data sets that will be included in the present study.

Ethics and dissemination: This study does not require ethics approval. Dissemination will include journal articles and presentations to policymakers, healthcare providers and researchers.

Systematic review registration: PROSPERO 2015: CRD42015024785.

INTRODUCTION

Depression is a leading cause of disability among women.¹ Although the 7–13% prevalence of major depression during pregnancy and postpartum^{2–5} is similar to rates among women during non-childbearing periods,^{4 6–10} perinatal depression is associated with adverse outcomes for the developing child, the mother–infant relationship and marital quality.^{11–13} Most women with depression in the perinatal period, however, do not receive adequate care.^{14–16} Thus, improving care of women with depression is a high clinical priority.¹⁷

Routine screening for depression, which involves the use of self-report questionnaires to identify patients with unrecognised depression who have not been otherwise identified as at risk for depression, has been proposed as a way to improve perinatal depression identification and management.^{18–20} However, screening for depression is controversial because it has not been shown to benefit patients in a well-designed and conducted randomised controlled trial, would lead to misidentification and overtreatment in some patients, and would consume large quantities of scarce resources without evidence of benefit.^{21–23} Indeed, recommendations, policy and implementation are inconsistent.

In 2010, the American College of Obstetricians and Gynecologists recommended that screening for depression be ‘strongly considered’ in pregnancy and the postpartum period; the report noted that there was not sufficient evidence to support a ‘firm recommendation’.¹⁸ Also in 2010, the American Academy of Pediatrics recommended that pediatricians screen new mothers for depression during well-child visits in the 6 months following birth.¹⁹ Neither of these recommendations, however, was based on a systematic review of the evidence. In the UK, the National Screening Committee determined in 2001, and again in 2010, that there is no evidence that postnatal screening would improve health outcomes.^{24 25} A 2014 UK National Institute for Health and Care Excellence guideline recommended that healthcare providers consider asking two questions about depression at several points during pregnancy and postpartum,²⁰ but this recommendation was not based on evidence from randomised controlled trials that this would improve health outcomes.

In Canada, postpartum depression screening is routine in Alberta,²⁶ but it is only considered for women at risk in Quebec²⁷ and not recommended in Ontario.²⁸ In 2013, the Canadian Task Force on Preventive Health Care recommended against screening for depression in primary care settings, including for women in pregnancy and postpartum. The Task Force noted that existing evidence likely overstates the accuracy of depression screening tools in practice and emphasised the need for better-quality estimates of screening tool accuracy.²⁹

Successfully screening for perinatal depression would require a screening tool with demonstrated accuracy. The 10-item Edinburgh Postnatal Depression Scale (EPDS) is the most commonly used screening tool in

pregnancy and postpartum.^{14 26 27} Most existing studies of the accuracy of depression screening tools, including the EPDS,³⁰ however, have been conducted in samples too small to precisely estimate accuracy. In a population with prevalence of 10%, for example, a screening study sample of 200 patients would only have approximately 20 patients to generate an estimate of sensitivity. Many studies have used data-driven methods to select ‘optimal’ cut-offs with these small samples and have selectively published accuracy results from high-performing cut-offs, but not other cut-offs, even when the other cut-offs are standard.^{31 32} Few studies have excluded women who are already being treated for depression and would not be screened in actual practice, thus exaggerating the estimated number of previously unidentified patients who would be detected by depression screening.³³ Ideally, primary studies that assess the diagnostic accuracy of the EPDS would include only women who have not already been identified as depressed, and results from all potentially relevant cut-off scores would be reported.

Meta-analyses can overcome small sample sizes in primary studies, but are unbiased only if all data from all relevant cut-off scores are included. Some meta-analyses^{34 35} have synthesised data to obtain a single accuracy estimate based on different cut-offs from different primary studies. In those meta-analyses, accuracy data from non-standard cut-offs have been included when primary studies only report optimally accurate cut-offs, but do not report less optimal accuracy results from standard cut-offs. Other meta-analyses^{30 36 37} have examined results from multiple cut-offs, but have similarly been limited to using published accuracy outcomes for each cut-off. The bias that can occur from synthesising results from selectively reported cut-offs is highlighted by some traditional meta-analyses,^{30 36} in which estimates of sensitivity actually increased as the cut-off score increased, suggesting that more patients would be detected when more severe symptoms were required for a positive screen. This is mathematically impossible if complete data are available. Beyond selective reporting, existing meta-analyses have not addressed limitations in primary studies from including already diagnosed and treated patients in study samples.³³

Individual patient data (IPD) meta-analysis can potentially address some of these problems. IPD meta-analysis involves using actual patient data obtained from researchers who conducted primary studies, rather than summary results from published or unpublished study reports.³⁸ The steps involved in conducting a systematic review with an IPD meta-analysis, in terms of defining a research question, establishing study inclusion and exclusion criteria, identifying and screening studies, and analysing data, is similar to a traditional systematic review and meta-analysis and diverges only in analysing individual level data rather than summary data.³⁹ When implemented effectively, IPD meta-analyses have particular benefits in that they can conduct analyses, including

some subgroup analyses, that cannot be addressed using study-level data available in original reports. In the context of evaluating the diagnostic accuracy of depression screening tools, IPD meta-analysis has a number of major advantages compared to traditional meta-analyses. First, IPD meta-analysis can address bias from the selective reporting in publications of only well-performing cut-off thresholds since accuracy can be evaluated across all relevant cut-off scores. Second, IPD meta-analysis allows the exclusion of already-treated patients, for whom the tool would not be used to screen for unidentified depression, in at least some primary data sets. Third, IPD meta-analysis with large numbers of patients and large numbers of depression cases allows the incorporation of factors that may influence screening accuracy (eg, age, pregnancy vs postpartum) and study variables (eg, study setting, risk of bias factors). Additionally, a large IPD meta-analysis would be able to generate estimates of the probability of having depression based on patient characteristics and actual EPDS scores, rather than classifying patients as simply negative or positive. This is an important consideration since a patient with a score of 0 on the EPDS, for example, would almost certainly have a lower likelihood of having depression than a patient with a substantially higher, but subthreshold, score of 10, although typically both would be classified as negative screens.

Potential downsides of IPD meta-analyses are that they are resource intensive and require substantial time to identify and obtain original data, clarify data-related issues with data providers, and generate a consistent data format across studies.^{38 39} Furthermore, if primary data sets cannot be obtained, results from the IPD meta-analysis can be biased. In some circumstances, the inability to obtain enough primary data may make it impossible to reasonably conduct an IPD meta-analysis.^{40 41} Currently, our team is conducting an IPD meta-analysis of the Patient Health Questionnaire depression screening tool, which is the first IPD meta-analysis of the diagnostic accuracy of a depression screening tool.⁴² In that study, which is still in process, as of 26 July 2015, we had identified 77 eligible primary data sets. Of these, 68 investigators had agreed to contribute their data sets, and 45 had already transferred data sets. This suggests that investigators are generally willing to provide primary data from studies of the diagnostic accuracy of depression screening tools for use in IPD meta-analyses. Preliminarily, we are aware that at least 70 eligible studies with the EPDS exist, and more may be identified with the complete search strategy.

Thus, the primary objective of this IPD meta-analysis is to determine the diagnostic accuracy of the EPDS to detect major depression among pregnant and postpartum women.

METHODS AND ANALYSIS

This systematic review has been funded by the Canadian Institutes for Health Research (Funding Reference

Number KRS-140994). The protocol has been registered in the PROSPERO prospective register of systematic reviews (2015:CRD42015024785), and any changes to the study protocol will be registered as amendments with PROSPERO.

The IPD meta-analysis has been designed and will be conducted in accordance with best-practice standards as elaborated in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy⁴³ and other key sources.^{38 39 44} Results will be reported in concordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.^{45 46} To conduct the meta-analysis, primary data sets will be sought that allow comparisons of EPDS scores to diagnoses of major depression.

Sources of evidence

The search strategy was developed by a medical librarian and peer-reviewed by another medical librarian. Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO on the OvidSP platform, and Web of Science (Web of Knowledge platform) will be searched. The Medline search strategy was validated by testing against already-identified publications from preliminary searches. The strategy was then adapted for PsycINFO and Web of Science. The search strategy will be limited to these databases based on research showing that adding other databases (eg, EMBASE) when the Medline search is highly sensitive does not identify additional eligible studies.⁴⁷ The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy⁴³ suggests combining concepts of the index test and the target conditions, but this was redundant for depression screening tools as these tests are limited to testing for depression. Thus, the search strategy for electronic databases was comprised of two concepts: the index test of interest and studies of screening accuracy. There are no published search hedges, or filters, designed specifically for mental health screening, but several key articles were consulted in developing search terms.^{48–50} Search strategies use a combination of subject headings, when available in the database, as well as keywords in the title, abstract, or anywhere else in the record. The search will be limited to the year 2000 forward since it is unlikely that many older data sets will be obtainable. See online supplementary file 1 for detailed information on searches. To supplement electronic searches, reference lists of all included publications and relevant reviews will be searched. In addition, a related articles search will be conducted for included papers indexed in Medline using the PubMed 'related articles' search feature. Researchers who have published on the topic will be contacted to obtain information about additional, unpublished studies. Search results will be initially uploaded into the citation management database RefWorks (RefWorks, RefWorks-COS, Bethesda, Maryland, USA), and the RefWorks duplicate check function will be used to identify citations retrieved from

multiple sources. Unique citations will then be uploaded into the systematic review program DistillerSR (Evidence Partners, Ottawa, Canada), and DistillerSR will be used to store and track search results and to track results of the review process.

To identify relevant data sets, articles published in any language will be reviewed. Data sets will be sought for inclusion if they compare results from the EPDS to Diagnostic and Statistical Manual (DSM) or International Classification of Diseases (ICD) criteria for major depression. ICD criteria are similar to DSM criteria and are generally used outside of North America. In DSM-5, there is not a separate diagnosis for depression during pregnancy or postpartum. Rather, cases of major depression are specified as having 'peripartum onset', which includes pregnancy and the postpartum period. Diagnoses must be based on a validated structured or semistructured interview (eg, Structured Clinical Interview for DSM,⁵¹ Composite International Diagnostic Interview⁵²) administered within 2 weeks of the administration of the depression screening tool, since major depression criteria are for symptoms in the last 2 weeks. Data sets where some patients were administered the screening tools within 2 weeks of the diagnostic interview and some patients were not will be included if the original data allow us to select patients administered the diagnostic interview and screening tools within the 2-week window. Data from studies where all patients are known to have psychiatric diagnoses will be excluded, with the exception of patients treated for substance and alcohol abuse, for whom depression screening may be considered. The coding manual for inclusion and exclusion decisions is shown in online supplementary file 2.

Two investigators will independently review titles and abstracts for eligibility. If either reviewer determines that a study may be eligible based on title or abstract review, then a full-text article review will be completed. Disagreement between reviewers after full-text review will be resolved by consensus, including a third investigator as necessary. Chance-corrected agreement between reviewers will be assessed with the Cohen's κ statistic. Translators will be consulted to evaluate titles/abstracts and articles for languages other than those for which team members are fluent (English, French, Spanish, Dutch, Greek). See online supplementary file 3 for preliminary PRISMA flow of studies figure.

Transfer of data and data set management

Authors of studies containing data sets that meet inclusion criteria will be contacted to invite them to contribute primary data for inclusion. Data will only be used from studies that received ethics approval and all data that are transferred will be properly de-identified prior to transfer. All individual patient data that are obtained will be cleaned and coded to make patient data as uniform as possible across data sets, then entered into a single STATA database (StataCorp, College Station,

Texas, USA). A preliminary codebook has been developed for coding data from original studies of the EPDS. For each study to be included in the data set, two investigators will independently determine the coding protocol, based on the codebook, with any discrepancies resolved by consensus. Actual data coding and transfer from original studies into the IPD database will be done by a supervised staff or trainee member of the team, and the resulting data sets will be compared to the original reports in order to identify any potential discrepancies. In addition to obtaining original patient-level data, data will also be extracted from the published articles of included studies. Any inconsistencies will be discussed with the original authors. Corrections will be made as necessary.

Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool⁵³ will be used to assess risk of bias factors in primary studies, and these factors will be included as study-level variables in analyses. QUADAS-2 incorporates assessments of risk of bias across four core domains: patient selection, the index test, the reference standard, and the flow and timing of assessments. Two reviewers will independently assess risk of bias with any discrepancies resolved by consensus.

Data analysis

Analyses will estimate sensitivity and specificity, which will be used to generate estimates of positive and negative predictive value, which are more useful clinically. A bivariate random-effects meta-analysis will be fit, estimated via Gauss-Hermite adaptive quadrature, as described in Riley *et al.*,⁵⁴ for the full range of plausible cut-off values, which for the EPDS will be considered cut-off scores from 9–15.³⁰ This approach appropriately models sensitivity and specificity simultaneously and also accounts for variation in within-study precision.⁵⁴ Data will be analysed using a random-effects model so that sensitivity and specificity are assumed to vary across studies. For each cut-off, separately, this model will provide an overall pooled sensitivity and specificity and an overall pooled diagnostic OR.⁵⁴

Analyses will be conducted that evaluate data from pregnancy and the postpartum period separately, as well as that combine data from all women in a single model. We anticipate that most primary studies include only one data point per woman. However, we expect that a small number of studies will provide data for more than one time point. To include all assessments, we will fit random effects models that account both for between study heterogeneity and between patient heterogeneity. Additionally, a subgroup analysis will be conducted that includes only data from countries listed as 'very high development' on the United Nation's Human Development Index⁵⁵ in order to capture studies from countries with healthcare resources and outcomes reasonably similar to those of Canada.

Heterogeneity will be quantified for each cut-off analysis by reporting the estimated variances of the random effects for sensitivity and specificity, as well as by estimating R . R is the ratio of the estimated SD of the pooled sensitivity from the random effects model to the estimated SD of the pooled sensitivity from the fixed effects model.⁵⁶ We will explore underlying reasons for heterogeneity using patient-level (eg, age) and study-level factors (eg, QUADAS-2 risk of bias ratings). In diagnostic accuracy this can easily be accomplished by including the factors or interaction terms in the random effects model described above.⁵⁴ These analyses take advantage of the richness of individual patient data. When analysed at the patient-level, accounting for correlation between patients from the same study, and for the correlation between sensitivity and specificity via the random effects model, they are more powerful to detect interactions and not vulnerable to ecological bias compared to traditional meta-analyses.^{57–61} It is not known if there are factors that influence the accuracy of depression screening tools in pregnancy or postpartum, but many risk factors have been identified for the presence of depressive symptoms. These may include lower education, lower income, lack of a partner, domestic violence, unintended pregnancy or complications in pregnancy, for example.^{62–63} Data relevant to established risk markers from primary study authors will be requested, although these are not likely to be consistently available. Many of these risk markers are associated with age below a threshold, and age below 23 or 25 has been used to stratify risk of depression.^{64–66} Thus, age (<25 vs ≥ 25 years) will be evaluated as a patient-level variable potentially associated with heterogeneity. It may be considered that, ideally, accuracy estimates would account for cultural or racial/ethnic differences between patients. However, existing studies have not been based on large enough samples to know if this would be the case. If there is enough group-specific data to adjust estimates in the IPD meta-analysis, we will do so. Study-level covariates, including risk of bias factors described in QUADAS-2,⁵³ may also be evaluated. QUADAS-2 factors include patient selection factors, blinding of reference standard to index test results, type of reference standard (eg, semistructured diagnostic interview; structured diagnostic interview), and timing of administration of index test and reference standard (eg, same day, delay of 1–7 days, delay of >7 days).

In addition to estimating sensitivity and specificity for each relevant cut-off, we will build predictive models that use the score on the screening questionnaire (combined model or separately for pregnancy and postpartum) and any other key factors that account for substantial heterogeneity to estimate the probability that a woman has major depression. The models will be evaluated in terms of their calibration (eg, slope of linear predictor; are average, low and high predictions correct?) and discrimination (eg, c-statistic; are low risk participants distinguished from high risk participants?).⁶⁷ Validation with

the same participants used to develop a model results in overly optimistic performance. Internal validation will be assessed via the bootstrap method, which is preferable to split sample validation approaches (eg, developing the model in half the sample and evaluating it in the other half).⁶⁸ Although there are advantages to external validation, given the wide range of study populations that will be used it would be unlikely that there would be another comparable data set large enough for validation. Thus, assessment of internal validity via bootstrapping will help develop an understanding of how the model will likely perform in a clinical setting, and by using the regression coefficients adjusted for optimism (ie, the shrinkage estimates), will maximise actual accuracy. Based on our pilot work, it is anticipated that missing data will be minimal for the variables of primary interest. Regardless, multiple imputation will be done using chained equations^{67–69} to impute data for both binary and continuous variables, considering study as a fixed effect in the imputation model.⁷⁰ This will allow imputation both for variables missing for entire studies as well those missing more sporadically.

In a sensitivity analysis, studies included in the IPD meta-analysis will be compared to eligible studies that do not provide data in terms of sensitivity and specificity using published summary data from the studies that do not provide data. A sensitivity analysis will also be conducted that includes aggregate summary estimates of sensitivity and specificity from the studies that do not provide individual patient data in the main meta-analysis.⁵⁴ If there are a large number of studies that do not contribute primary data, this analysis may become the primary analysis.

ETHICS AND DISSEMINATION

This IPD meta-analysis does not require ethics approval, although only individual studies that obtained ethical clearance and informed consent will be included. The reasons that the IPD meta-analysis does not require ethics review is that the objectives of the IPD meta-analysis are consistent with the objectives of the primary studies, which already received ethics approval, and only anonymised data will be provided by the investigators of the original studies.

The main outcomes of the IPD meta-analysis reflect knowledge that will influence policy and clinical practice. The proposed project is a result of a collaborative process, which involved research team members and knowledge users at the national, provincial, regional and local levels of Canada. Currently, as described in the 2013 Canadian Task Force on Preventive Health Care guideline on screening for depression, policymakers must make decisions about screening for depression in the absence of well-designed and conducted clinical trials and with primary evidence on depression screening tool accuracy that does not appear to reflect what would likely occur in practice. Regional and local

planners and individual clinicians are faced with the same problem, as well as inconsistent guidance on whether screening for depression should be included in clinical care. The present IPD meta-analysis will provide high-quality evidence on screening tool accuracy to these stakeholders. It will also increase the ability of researchers to incorporate better evidence-informed and potentially individualised screening protocols into trials of depression screening programmes.

At the national level, the key knowledge user is Dr Marcello Tonelli, who is the Chair of the Canadian Task Force on Preventive Health Care. At the provincial, regional and local levels, the research team has partnered with a key provincial level policymaker with an important role in perinatal mental health policy (Dr Nicholas Mitchell, Alberta, Canada), an advisor with a provincial institute that provides policy-relevant expertise to health authorities on a range of topics, including perinatal health (Dr Liane Comeau, Quebec, Canada), a key regional planning officer with experience in perinatal policy implementation (Ms Joy Schinazi), and a psychiatrist whose practice focuses on mental illness in pregnancy and postpartum (Dr Simone Vigod). These partnerships will facilitate successful dissemination of the results of this IPD meta-analysis to provincial health agencies, non-governmental organisations, clinical programmes and university training programmes. Internationally, team members will work to disseminate findings to national guideline-making and policy-making bodies outside of Canada, such as the National Institute for Health and Care Excellence and the National Screening Committee in the UK, for instance.

Strategies for effective dissemination and specific outputs will be based on research showing how to best tailor research outputs to different user groups,^{71–76} including research on improving the usefulness of reports of systematic review and meta-analyses for health-care managers and policymakers.^{74 76} Dissemination will include publication of results in high-impact medical journals with open access, as well as presentations in seminars and symposia to policymakers, healthcare providers and researchers at national and international conferences.

Traditional methods, however, such as journal articles and conference presentations are typically not highly effective in disseminating clinical research to healthcare providers, including physicians.⁷⁷ To increase uptake of the findings from this study, a brief video and podcast will be created that is similar to the Cochrane Library podcasts (<http://www.cochrane.org/podcasts/>). These will summarise the key findings of these meta-analyses and their implications for practice. They will also provide an introduction to an online calculation tool that will be created and made freely available to estimate the probability that a given patient has major depression based on depression screening results and patient characteristics. An example of a tool that is based on

robust research evidence and effectively disseminated is the FRAX Fracture Risk Assessment Tool (<http://www.shef.ac.uk/FRAX/index.aspx>). The tool that will be made from the results of this study will be modelled on this tool and presented similarly in an easy-to-use fashion with tablet and app versions. In addition, simpler nomogram-based presentations, which are user-friendly graphical depictions of positive and negative predictive value by prevalence, will be generated and made available.

Author affiliations

- ¹Departments of Psychiatry, McGill University, Montreal, Quebec, Canada
- ²Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada
- ³Department of Medicine, McGill University, Montreal, Quebec, Canada
- ⁴Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada
- ⁵Department of Psychology, McGill University, Montreal, Quebec, Canada
- ⁶School of Nursing, McGill University, Montreal, Quebec, Canada
- ⁷Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada
- ⁸Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, Quebec, Canada
- ⁹Library, McGill University, Montreal, Quebec, Canada
- ¹⁰Department of Clinical, Neuro, Developmental Psychology, VU-University, Amsterdam, The Netherlands
- ¹¹Psychological Medicine and Health Services Research, Department of Health Sciences, Hull York Medical School University of York, York, UK
- ¹²Department of Medicine, Department of Health Research and Policy, Department of Statistics, Stanford Prevention Research Center, Stanford School of Medicine, Stanford University School of Humanities and Sciences, Stanford, California, USA
- ¹³Departments of Psychiatry and Community Health Sciences, University of Calgary, Calgary, Alberta, Canada
- ¹⁴Department of Mathematics and Statistics, McGill University, Montreal, Quebec, Canada
- ¹⁵Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
- ¹⁶Department of Medicine, University of Calgary, Calgary, Alberta, Canada
- ¹⁷Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada
- ¹⁸Institut national de santé publique du Québec and Université de Montréal, Montreal, Quebec, Canada
- ¹⁹Public Health Department, Laval, Quebec, Canada
- ²⁰Department of Psychiatry, Women's College Hospital and Women's College Research Institute, University of Toronto, Toronto, Ontario, Canada

Twitter Follow Simon Gilbody at @SimonGilbody

Contributors BDT, AB, LAK, BL, KER, MA, PC, SG, JPAI, DMcM, SBP, IS, RJS, RCZ, MT, NM, LC, JS and SV contributed to the conception and design of the systematic review and meta-analysis. BDT, AB, LAK, BL, KER, and MA will be involved in acquisition of data. BDT, AB, and BL will analyse the data. BDT, AB, LAK, BL, KER, MA, PC, SG, JPAI, DMcM, SBP, IS, RJS, RCZ, MT, NM, LC, JS, and SV will interpret the results. BDT and AB drafted this protocol. All authors provided critical revisions of the protocol and approved submission of the final manuscript. BDT is the guarantor.

Funding This research was supported by a grant from the Canadian Institutes for Health Research (CIHR; Funding Reference Number KRS-140994; PI Thombs). BDT receives support from an Investigator Award from the Arthritis Society. AB is supported by a Fonds de recherche santé Québec (FRSQ) researcher salary award. BL is supported by a FRSQ doctoral award. MA is supported by a CIHR Frederick Banting and Charles Best Canadian Graduate Scholarships—Master's Award. SV receives support from a New Investigator Fellowship from the Ontario Mental Health Foundation. No funding body had any input into any aspect of this protocol.

Competing interests SV received an honorarium from MDH consulting in 2012 for the development of continuing medical education activities related to perinatal mental health.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

Data sharing statement Data related to this research article, once results are published, can be obtained by contacting the corresponding author.

Open Access 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/>

REFERENCES

- Kessler RC. Epidemiology of women and depression. *J Affect Disord* 2003;74:5–13.
- Gaynes BN, Gavin N, Meltzer-Brody S, et al. Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes. Evidence Report/Technology Assessment No. 119. (Prepared by the RTI-University of North Carolina Evidence-based Practice Center, under Contract No. 290-02-0016.) AHRQ Publication No. 05-E006-2. Rockville, MD: Agency for Healthcare Research and Quality. February 2005.
- O'Hara MW, Swain AM. Rates and risk of postpartum depression—a meta-analysis. *Int Rev Psychiatr* 1996;8:37–54.
- Gavin NI, Gaynes BN, Lohr KN, et al. Perinatal depression: a systematic review of prevalence and incidence. *Obstet Gynecol* 2005;106:1071–83.
- Bennett HA, Einarson A, Taddio A, et al. Prevalence of depression during pregnancy: systematic review. *Obstet Gynecol* 2004;103:698–709.
- Cooper PJ, Campbell EA, Day A, et al. Non-psychotic psychiatric disorder after childbirth. A prospective study of prevalence, incidence, course and nature. *Br J Psychiatry* 1988;152:799–806.
- O'Hara MW, Zekoski EM, Philipps LH, et al. Controlled prospective study of postpartum mood disorders: comparison of childbearing and nonchildbearing women. *J Abnorm Psychol* 1990;99:3–15.
- Cox JL, Murray D, Chapman G. A controlled study of the onset, duration and prevalence of postnatal depression. *Br J Psychiatry* 1993;163:27–31.
- Stewart DE. Clinical practice. Depression during pregnancy. *N Engl J Med* 2011;365:1605–11.
- Vesga-Lopez 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.
- Zelkowitz P, Milet TH. The course of postpartum psychiatric disorders in women and their partners. *J Nerv Ment Dis* 2001;189:575–82.
- Zelkowitz P, Milet TH. Postpartum psychiatric disorders: their relationship to psychological adjustment and marital satisfaction in the spouses. *J Abnorm Psychol* 1996;105:281–5.
- Whitley R, Kirmayer LJ. Perceived stigmatisation of young mothers: an exploratory study of psychological and social experience. *Soc Sci Med* 2008;66:339–48.
- Howard LM, Molyneaux E, Dennis C, et al. Non-psychotic mental disorders in the perinatal period. *Lancet* 2014;384:1775–88.
- Duhoux A, Fournier L, Gauvin L, et al. What is the association between quality of treatment for depression and patient outcomes? A cohort study of adults consulting in primary care. *J Affect Disord* 2013;151:265–74.
- Duhoux A, Fournier L, Nguyen CT, et al. Guideline concordance of treatment for depressive disorders in Canada. *Soc Psychiatry Psychiatr Epidemiol* 2009;44:385–92.
- Mental Health Commission of Canada. *Changing directions, changing lives: the mental health strategy for Canada*. Calgary, Alberta, Canada: Mental Health Commission of Canada, 2012.
- American College of Obstetricians and Gynecologists. Committee on Obstetric Practice. Committee opinion no. 453: Screening for depression during and after pregnancy. *Obstet Gynecol* 2010;115:394–5.
- Earls MF, Committee on Psychosocial Aspects of Child and Family Health American Academy of Pediatrics. Incorporating recognition and management of perinatal and postpartum depression into pediatric practice. *Pediatrics* 2010;126:1032–9.
- National Institute for Health and Care Excellence. *Antenatal and postnatal mental health: clinical management and service guidance*. London, UK: National Institute for Health and Care Excellence, 2014.
- Thombs BD, Arthurs E, Coronado-Montoya S, et al. Depression screening and patient outcomes in pregnancy or postpartum: a systematic review. *J Psychosom Res* 2014;76:433–46.
- Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? *BMJ* 2014; 348:g1253.
- Thombs BD, Ziegelstein RC, Roseman M, et al. There are no randomized controlled trials that support the United States Preventive Services Task Force guideline on screening for depression in primary care: a systematic review. *BMC Med* 2014;12:13.
- Shakespeare J. *Evaluation of screening for postnatal depression against the NSC handbook criteria*. UK: National Screening Committee, 2001.
- Hill C. *An evaluation of screening for postnatal depression against NSC criteria*. UK: National Screening Committee, 2010.
- Health Information Standards Committee for Alberta. *Alberta postpartum depression data set*. Edmonton: Government of Alberta, 2009.
- Direction générale des services de sante et médecine universitaire. *Politique de périnatalité 2008–2018*. Quebec: Sante et services sociaux Québec, 2012.
- Registered Nurses' Association of Ontario. *Interventions for postpartum depression*. ON, Canada: Registered Nurses' Association of Ontario, 2005.
- Joffres M, Jaramillo A, Dickinson J, et al. Recommendations on screening for depression in adults. *CMAJ* 2013;185:775–82.
- Hewitt C, Gilbody S, 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;13:1–145, 147–230.
- Thombs BD, de Jonge P, Coyne JC, et al. Depression screening and patient outcomes in cardiovascular care: a systematic review. *JAMA* 2008;300:2161–71.
- Meijer A, Roseman M, Milette K, et al. Depression screening and patient outcomes in cancer: a systematic review. *PLoS ONE* 2011;6: e27181.
- Meader N, Mitchell AJ, Chew-Graham C, et al. Case identification of depression in patients with chronic physical health problems: a diagnostic accuracy meta-analysis of 113 studies. *Br J Gen Pract* 2011;61:e808–20.
- Meader N, Moe-Byrne T, Llewellyn A, et al. Screening for poststroke major depression: a meta-analysis of diagnostic validity studies. *J Neurol Neurosurg Psychiatry* 2014;85:198–206.
- Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ* 2012;184:e191–6.
- Brennan C, Worrall-Davies A, McMillan D, et al. The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. *J Psychosom Res* 2010;69:371–8.
- 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.
- Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; 340:c221.
- Stewart LA, Tierney JF, Clarke M. Chapter 18: reviews of individual patient data. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions version 5.1.0*. Chichester, UK: Wiley-Blackwell; The Cochrane Collaboration, 2011:18.1–18.9.
- Jaspers GJ, Degraeuwe PL. A failed attempt to conduct an individual patient data meta-analysis. *Syst Rev* 2014;3:97.
- Parish E, Bloom T, Godlee F. Statins for people at low risk. *BMJ* 2015;351:h3908.
- Thombs BD, Benedetti A, Kloda LA, et al. The diagnostic accuracy of the patient health questionnaire-2 (PHQ-2), patient health questionnaire-8 (PHQ-8), and patient health questionnaire-9 (PHQ-9) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. *Syst Rev* 2014;3:124.
- Deeks JJ, Bossuyt PM, Gatsonis C. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0.0*. The Cochrane Collaboration, 2009.
- Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions version 5.1.0*. Chichester, UK: Wiley-Blackwell; The Cochrane Collaboration, 2011.

45. Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
46. Liberati A, Altman DG, Tetzlaff J, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6:e1000100.
47. Sampson M, Barrowman NJ, Moher D, *et al.* Should meta-analysts search Embase in addition to Medline? *J Clin Epidemiol* 2003;56:943–55.
48. Mann R, Gilbody SM. Should methodological filters for diagnostic test accuracy studies be used in systematic reviews of psychometric instruments? A case study involving screening for postnatal depression. *Syst Rev* 2012;1:9.
49. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001;8:391–7.
50. Wong SS, Wilczynski NL, Haynes RB, *et al.* Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. *AMIA Annu Symp Proc* 2003:728–32.
51. First MB, Spitzer RL, Gibbon M, *et al.* *Structured clinical interview for DSM-IV Axis I Disorders—Patient Edition (SCID-I/P, Version 2.0)*. New York: Biometrics Research Department, New York State Psychiatric Institute, 1996.
52. Wittchen HU. Reliability and validity studies of the WHO—Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res* 1994;28:57–84.
53. 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.
54. 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.
55. United Nations Development Programme. *Human development report 2013. The Rise of the South: Human Progress in a Diverse World*. New York, NY: United Nations Development Programme, 2013.
56. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539–58.
57. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998;54:638–45.
58. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2007;26:2982–99.
59. Berlin JA, Santanna J, Schmid CH, *et al.* Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002;21:371–87.
60. Lambert PC, Sutton AJ, Abrams KR, *et al.* A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002;55:86–94.
61. Schmid CH, Stark PC, Berlin JA, *et al.* Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol* 2004;57:683–97.
62. Robertson E, Grace S, Wallington T, *et al.* Antenatal risk factors for postpartum depression: a synthesis of recent literature. *Gen Hosp Psychiatry* 2004;26:289–95.
63. Lancaster CA, Gold KJ, Flynn HA, *et al.* Risk factors for depressive symptoms during pregnancy: a systematic review. *Am J Obstet Gynecol* 2010;202:5–14.
64. 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 Womens Ment Health* 2005;8:97–104.
65. Leigh B, Milgrom J. Risk factors for antenatal depression, postnatal depression and parenting stress. *BMC Psychiatry* 2008;8:24.
66. Rich-Edwards JW, Kleinman K, Abrams A, *et al.* Sociodemographic predictors of antenatal and postpartum depressive symptoms among women in a medical group practice. *J Epidemiol Community Health* 2006;60:221–7.
67. Steyerberg EW. *Clinical prediction models*. New York: Springer, 2009.
68. Steyerberg EW, Harrell FE, Borsboom GJ, *et al.* Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
69. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med* 2009;28:1982–98.
70. Riley RD, Takwoingi Y, Trikalinos T, *et al.* Meta-analysis of test accuracy studies with multiple and missing thresholds: a multivariate-normal model. *J Biomed Biostat* 2014;5:196.
71. Dobbins M, DeCorby K, Twiddy T. A knowledge transfer strategy for public health decision makers. *Worldviews Evid Based Nurs* 2004;1:120–8.
72. Graham ID, Logan J, Harrison MB, *et al.* Lost in knowledge translation: time for a map? *J Contin Educ Health Prof* 2006;26:13–24.
73. Haines A, Kuruville S, Borchert M. Bridging the implementation gap between knowledge and action for health. *Bull World Health Organ* 2004;82:724–31.
74. Dobbins M, Thomas H, O'Brien MA, *et al.* Use of systematic reviews in the development of new provincial public health policies in Ontario. *Int J Technol Assess Health Care* 2004;20:399–404.
75. Dobbins M, Jack S, Thomas H, *et al.* Public health decision-makers' informational needs and preferences for receiving research evidence. *Worldviews Evid Based Nurs* 2007;4:156–63.
76. Lavis J, Davies H, Oxman A, *et al.* Towards systematic reviews that inform health care management and policy-making. *J Health Serv Res Policy* 2005;10(Suppl 1):35–48.
77. Beaulieu MD, Proulx M, Jobin G, *et al.* When is knowledge ripe for primary care? An exploratory study on the meaning of evidence. *Eval Health Prof* 2008;31:22–42.