Systematic Review Protocol: A meta-analysis for the diagnostic properties of the Whooley Questionnaire to identify depression in perinatal women

Background (800 words max)

Depression is the leading cause of disability worldwide, it is a major contributor to the global burden of disease, affecting a variety of populations (1). Depression experienced during pregnancy, also known as perinatal depression, is common and can affect up to 20% of pregnant mothers (2). Perinatal depression can lead to a range of adverse health outcomes for the mothers and the development of their children. Depression during pregnancy can lead to somatic problems including headaches and gastrointestinal pain, poorer self-report health and functioning and increased risk of smoking or alcohol abuse (3). Difficulties during parenting leading from perinatal depression can include lower interaction and sensitivity between mother and infant. Perinatal depression has been shown to lead to difficulties in later years within the offspring including emotionally, socially and developing at school (4-6).

Despite its high prevalence and large impact on outcomes, less than half of pregnant women suffering from depression are identified within healthcare (7). Attitudes towards perinatal depression case-finding among clinicians is positive, there is a need for support and strategies to identify perinatal depression within routine practice (8). The Gold standard approach for diagnosing depression is using the psychiatric diagnostic criteria (SCID), the SCID requires a clinician using standardised interviewing to establish a diagnosis of depression and is not suitable for screening purposes given its requirements for a clinician and time-consuming interviewing.

A case-finding tool that has been identified as promising for use in identifying cases of perinatal depression are two brief questions, referred to as the Whooley questionnaire that can be used by healthcare professionals within a clinical setting. The Whooley features two questions with a third question if the participant replies yes to any of the 2 previous questions (9). Meta-analysis of the Whooley questionnaire among general populations in community, primary care and secondary care settings suggested high sensitivity, modest specificity, high diagnostic odds ratio (OR) and high likelihood ratios compared to Gold standard diagnosis. Indicating that the Whooley questionnaire to have acceptable diagnostic properties for use in a clinical setting (10). Within the previous meta-analysis, only one study was conducted in a perinatal population. Since then, a scoping search has shown further studies have emerged evaluating the diagnostic properties of the Whooley questionnaire in perinatal populations, although there has not been a pooling of this data to give an evaluation of the Whooley questionnaire across different studies to provide best possible available evidence.

Synthesis of evidence for the diagnostic properties of the Whooley questionnaire in perinatal populations would address the important gap in evidence. Providing guidance for the appropriateness of the questionnaire in clinical practice for identifying those with perinatal depression within healthcare. Facilitating positive cases to be offered treatments and avoid associated adverse clinical outcomes.

Review Question

To establish what the diagnostic properties of the Whooley questionnaire are in perinatal populations.

Method

1. Searches

Searches of online databases will commence on 21st October 2019. The search strategy has been developed based on the PICOS for this meta-analysis **(Table 1.)**

Table 1. PICOS for search strategy and study selection criteria

Population	Perinatal mothers, or if not identified as perinatal within the study; females between 20th week of gestation to 4 weeks after birth, no limitation of setting
Instrument	Either the 2-item or 3-item version of the Whooley questionnaire
Comparison	Gold standard diagnostic interview
Outcome	Sufficient data to extract a two-by-two table
Study design	No restriction

We identified search terms taken from a previously published meta-analysis on the diagnostic properties of the Whooley questionnaire (10). We further added search terms for perinatal populations using population specific search terms taken from a Cochrane review for psychosocial assessment for reducing perinatal mental health morbidity (11). An example full list of search terms for the MEDLINE database can be found in **Appendix 1**.

Databases searched electronically will be MEDLINE, British Nursing Index, EMBASE, Cumulative Index to Nursing & Allied Health (CINAHL Plus), PsycINFO, Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR).

Our search strategy shall include a time limitation within all databases. Searches will be conducted from 1994 to the date of search, the year in which the Whooley questionnaire was first developed. We shall conduct a forward citation search of the original Whooley questionnaire paper using the Web of Science database additional papers. We examined the reference lists and forward citations of all included studies.

2. Data Selection

Our study selection inclusion and exclusion criteria shall be based on the PICOS identified in **Table 1**. Retrieved articles from the search shall be downloaded on to Endnote where articles shall be scanned in two stages: title and abstract then full article review. Inclusion and exclusion criteria that articles shall be scanned against is located in **Table 2**. Study selection shall be conducted by two reviewers, with both reviewers reviewing each retrieved article independently. Differences in assessment or any disagreements over the eligibility of particular studies were resolved by discussion, and in cases of disagreement, the third reviewers would be consulted. Rationale for exclusion of studies will be recorded at both stages of study selection. Piloting of the data selection process shall take place prior to the full study selection, with 10 articles randomly selected from the search retrieved. Adjustments to the inclusion and exclusion criteria and study selection process may be made following from the piloting.

Table 2. inclusion and exclusion criteria for study selection

Inclusion Criteria	**Exclusion Criteria**
• Perinatal mothers, or if not identified as perinatal within the study; females between 20th week of gestation to 4 weeks after birth • Studies that exclude people already known to be depressed • Data to extract a two-by-two table of either the 2-item or 3-item version of the Whooley questionnaire compared to a gold standard diagnostic interview	• Population with mixture of perinatal mothers and others, where data for perinatal mothers only cannot be extracted

3. Data Extraction

Data exaction shall be conducted by both review independently using a standardised extraction form. The extraction form is based on a previously used form in a systematic review on the diagnostic characteristics of the Whooley questionnaire (10). The extraction form includes descriptive statistics of the studies' sample, design and setting. Descriptive data on the Whooley questionnaire (mode of administration, who administered, language, 2-item or 3-item), description of the gold standard diagnosis approach (interview technique, and diagnosis criteria) and the 2-by-2 contingency table for Whooley questionnaire against the gold standard diagnosis approach. Differences in assessment or any disagreements for data extraction of particular studies were resolved by discussion, and in cases of disagreement, the third reviewers would be consulted.

4. Risk of bias assessment

Risk of bias will be assessed using an adapted version of the Quality Assessment of Diagnostic Accuracy Studies version 2 tool (QUADAS-II) (12). This tool includes 4-domains based on risk of biases within diagnostic characteristics studies; patient selection, index test, reference standard and flow and timing. The adapted version of the QUADAS-II will be changed to remove an item of threshold of the index test, as the Whooley questionnaire is dichotomous in scoring. Results of the risk of bias shall be presented within each domain of risk of bias rather than an overall score for each study. Piloting and further adaption of the risk of bias assessment, if necessary, shall be conducted prior to the full review as recommended by authors of the QUADAS-II (12). For the item 'was there an appropriate interval between index testing and reference standard?' we took intervals less than 2 weeks as acceptable, following previous studies evaluation on the diagnosis of depression (10). The likelihood of correct classification of the reference standard was evaluated depend on the training of the interviewer who conducted the gold standard diagnosis (10). Risk of bias was evaluated independently by two reviewers, differences in assessment or any disagreements of particular studies were resolved by discussion, and in cases of disagreement, the third reviewers would be consulted.

5. Strategy for data synthesis

Data from each study shall be used to construct a 2-by-2 contingency table with true positive, true negative, false positive and false negative values. A bivariate mixed-effect meta-analysis shall be performed to give pooled estimations of sensitivity, specificity, likelihood ratios and diagnosis odds ratios (DOR). The mixed regression accounts for precision within studies on sensitivity and specificity and an estimation of between study differences. Heterogeneity shall be assessed using the I^2 statistic for all pooled diagnostic values separately, a value of 0% indicates no observed heterogeneity, and values greater than 50% will be considered substantial heterogeneity (13). Cases of substantial heterogeneity shall be investigated for risk of bias and study characteristics as covariates within a meta-regression using a logit regression of DOR. Publication bias shall be estimated using a funnel plot of the DOR. As the Whooley Questionnaire is dichotomously scored (positive or negative) there is no risk to threshold effect within our analysis of heterogeneity or publication bias.

A priori variables for detecting causes of heterogeneity include; study setting (hospital based - primary care/community setting), original or translated version of the Whooley questionnaire, prevalence of depression, prevalence of major depression, each domain of risk of bias assessment.

All analysis shall be conducted using STATA v14.0, using the midas command.

6. Analysis of subgroups or subsets

In addition to the subgroup analysis that maybe conducted in connection to variables that cause heterogeneity between studies, we shall also conduct a subgroup comparing the diagnostic characteristics of the 2-item and 3-item Whooley Questionnaire.

7. Ethics & Dissemination plans

Ethics is not required for this study, given that this is a protocol for a systematic review, which utilizes published data. The results of the review would be widely disseminated locally, nationally and internationally. A paper would be submitted to a leading peer-review journal in this field, reporting of the study will adhere to the PRISMA-DTA statement on the reporting of diagnostic accuracy studies(14). The findings shall also be presented at a relevant international conference.

8. References

1. Friedrich MJ. Depression Is the Leading Cause of Disability Around the World. JAMA. 2017; 317(15): 1517-.
2. Woody C, Ferrari A, Siskind D, Whiteford H, Harris M. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. Journal of Affective Disorders. 2017; 219: 86-92.
3. Marcus SM. Depression during pregnancy: rates, risks and consequences. Journal of Population Therapeutics and Clinical Pharmacology. 2009; 16(1).
4. Murray L, Arteche A, Fearon P, Halligan S, Goodyer I, Cooper P. Maternal postnatal depression and the development of depression in offspring up to 16 years of age. Journal of the American Academy of Child & Adolescent Psychiatry. 2011; 50(5): 460-70.
5. Velders FP, Dieleman G, Henrichs J, Jaddoe VW, Hofman A, Verhulst FC, et al. Prenatal and postnatal psychological symptoms of parents and family functioning: the impact on child emotional and behavioural problems. European child & adolescent psychiatry. 2011; 20(7): 341-50.
6. Kersten-Alvarez LE, Hosman CM, Riksen-Walraven JM, van Doesum KT, Smeekens S, Hoefnagels C. Early school outcomes for children of postpartum depressed mothers: comparison with a community sample. Child Psychiatry & Human Development. 2012; 43(2): 201-18.
7. Yonkers KA, Wisner KL, Stewart DE, Oberlander TF, Dell DL, Stotland N, et al. The management of depression during pregnancy: a report from the American Psychiatric Association and the American College of Obstetricians and Gynecologists. General hospital psychiatry. 2009; 31(5): 403-13.
8. LaRocco-Cockburn A, Melville J, Bell M, Katon W. Depression screening attitudes and practices among obstetrician–gynecologists. Obstetrics & Gynecology. 2003; 101(5): 892-8.
9. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression: Two questions are as good as many. Journal of general internal medicine. 1997; 12(7): 439-45.
10. Bosanquet K, Bailey D, Gilbody S, Harden M, Manea L, Nutbrown S, et al. Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. BMJ open. 2015; 5(12): e008913.
11. Austin MP, Priest SR, Sullivan EA. Antenatal psychosocial assessment for reducing perinatal mental health morbidity. Cochrane Database of Systematic Reviews. 2008; (4).
12. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011; 155(8): 529-36.
13. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in medicine. 2002; 21(11): 1539-58.
14. McInnes MD, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. Jama. 2018; 319(4): 388-96.