BMJ Open

Identification of depression in women during pregnancy and the early postnatal period using the Whooley questions and the Edinburgh Postnatal Depression Scale: protocol for the Born and Bred in Yorkshire - PeriNatal Depression Diagnostic Accuracy (BaBY PaNDA) study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011223
Article Type:	Protocol
Date Submitted by the Author:	20-Jan-2016
Complete List of Authors:	Littlewood, Elizabeth; The University of York, Health Sciences Ali, Shehzad; University of York, Department of Health Sciences Ansell, Pat; University of York, Department of Health Sciences Dyson, Lisa; University of York, York Trials Unit Gascoyne, Samantha; University of York, Department of Health Sciences Hewitt, Catherine; The University of York, Health Sciences Keding, Ada; University of York, Health Sciences Mann, Rachel; University of York, Social Policy Research Unit McMillan, Dean; University of York, Department of Health Sciences Morgan, Deborah Swan, Kelly; University of York, Department of Health Sciences Waterhouse, Bev; Calderdale and Huddersfield NHS Foundation Trust, Children, Women and Families Division Gilbody, Simon; The University of York, Department of Health Sciences
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Health services research
Keywords:	diagnostic accuracy, Whooley questions, perinatal depression, Edinburgh Postnatal Depression Scale, screening

TITLE: Identification of depression in women during pregnancy and the early postnatal period using the Whooley questions and the Edinburgh Postnatal Depression Scale: protocol for the Born and Bred in Yorkshire – PeriNatal Depression Diagnostic Accuracy (BaBY PaNDA) study

Elizabeth Littlewood[1]*, Shehzad Ali[1], Pat Ansell[2], Lisa Dyson[3], Samantha Gascoyne[1], Catherine Hewitt[3], Ada Keding[3], Rachel Mann[4], Dean McMillan[1,5], Deborah Morgan[6], Kelly Swan[2], Bev Waterhouse[7], Simon Gilbody[1,5] on behalf of the BaBY PaNDA study team.

Author affiliations

[1]Mental Health and Addiction Research Group, Department of Health Sciences, University of York, York, YO10 5DD, UK

[2]Epidemiology and Cancer Statistics Research Group, Department of Health Sciences, University of York, York, YO10 5DD, UK

[3]York Trials Unit, Department of Health Sciences, University of York, York, YO10 5DD, UK

[4]Social Policy Research Unit, University of York, York, YO10 5DD, UK

[5]Hull York Medical School, University of York, York, YO10 5DD, UK

[6]PNI-UK, Ashbourne, Derbyshire, DE6 5NB, UK

[7]Children, Women & Families Division, Calderdale and Huddersfield NHS Foundation Trust, Calderdale Royal Hospital, Salterhebble, Halifax, HX3 0PW, UK

***Corresponding author**

Elizabeth Littlewood

Mental Health and Addiction Research Group

Department of Health Sciences

ARRC Building, Area 4

University of York

York

YO10 5DD, UK

Email: liz.littlewood@york.ac.uk

Tel: +44 (0)1904 321828

Key words: diagnostic accuracy; Whooley questions; perinatal depression; Edinburgh Postnatal Depression Scale, screening

Word count: 5,170; **Abstract; Tables: 1**

ABSTRACT

Introduction

Perinatal depression is well recognised as a mental health condition but less than 50% of cases are identified by healthcare professionals in routine clinical practice. The Edinburgh Postnatal Depression Scale (EPDS) is often used to detect postnatal depression in maternity and child services. The National Institute of Health and Care Excellence (NICE) recommends two 'ultra-brief' case-finding questions (the Whooley questions) to aid identification of depression during the perinatal period, but this recommendation was made in the absence of any validation studies in a perinatal population. There is limited research on the acceptability of these depression case-finding instruments and on the cost-effectiveness of routine screening for perinatal depression.

Methods and analysis

The diagnostic accuracy of the Whooley questions and the EPDS will be determined against a diagnostic gold standard during pregnancy (around 20 weeks) and the early postnatal period (around 3-4 months post-birth) in a sample of 379 women. Secondary outcome measures will assess psychological comorbidity, health related quality of life and resource utilisation. Women will be followed up 12 months after birth. The sensitivity, specificity and predictive values of the depression case-finding instruments will be calculated against the diagnostic gold standard at 20 weeks pregnancy and 3-4 months post-birth. Acceptability of the Whooley questions and the EPDS to women and healthcare professionals will involve in-depth qualitative interviews and completion of an acceptability survey. A decision analytic model will be adapted to determine the cost-effectiveness of routine screening for perinatal depression.

Ethics and dissemination

This study is considered low risk for participants. Robust protocols will deal with cases where risk of depression, self-harm or suicide is identified. The protocol received favourable ethical opinion from the North East – York Research Ethics Committee (reference: 11/NE/0022). The study findings will be published in peer-reviewed journals and presented at relevant conferences.

Strengths and limitations of this study

- This study will fill an important evidence gap regarding the diagnostic utility of depression case-finding instruments for the identification of perinatal depression.
- The study findings will be informed by qualitative interviews with women and healthcare professionals regarding their acceptability of depression case-finding instruments administered during the perinatal period.
- An existing decision analytic model will be updated with current diagnostic accuracy estimates of two depression case-finding instruments, providing an up-to-date estimate of the cost-effectiveness of a perinatal depression screening strategy.
- The study findings will inform policy decisions on the implementation of screening and case-finding strategies for the identification of perinatal depression.
- The study spans four NHS trusts which may implement differing policies regarding the identification of and referral processes for perinatal depression during the perinatal period.

INTRODUCTION

Depression accounts for the greatest burden of disease of all mental health problems and is estimated to become the second largest cause of global disability by 2020 [1]. It is well recognised that perinatal depression, that is depression experienced during pregnancy and/or the postnatal period (up to one year after birth), is an important category of depression in its own right, with specific guidance provided on the identification and clinical management of the condition[2, 3].

Prevalence rates of perinatal depression vary. Estimates indicate that approximately 7.4%-20% of women experience depression at some stage during pregnancy[4-6] with depression during the postnatal period affecting up to 22% of women[6]. Perinatal depression is associated with a range of adverse outcomes. Evidence suggests an association between antenatal depression (depression experienced during pregnancy) and adverse neonatal outcomes, poor self-reported health, substance abuse and alcohol abuse, and poor usage of antenatal care services[5]. Postnatal depression has been shown to have a substantial impact on the mother and her partner[7], mother-baby interactions[8], the family[9] and on the longer-term emotional and cognitive development of the baby[10], particularly when depression occurs in the first year of life[11].

Although perinatal depression is well recognised as a mental health condition, it often goes undetected; with healthcare professionals detecting less than 50% of cases in routine clinical practice[12]. The National Service Framework (NSF) states that local protocols should be in place for the management of postnatal depression[13], promoting the use of case-finding or screening strategies to aid identification of depression during the perinatal period. This has led to the routine or ad-hoc administration of self-report measures such as the Edinburgh Postnatal Depression Scale (EPDS)[14]. Screening or case-finding strategies such as those advocated by the NSF have since come under scrutiny[15] and have been criticised on a number of factors. Criticisms of the proposed strategies are based on the ethics of mass screening, concerns regarding the psychometric properties of available screening or case-finding instruments (such as the EPDS), the acceptability of such screening or case-finding strategies to patients and healthcare professionals, and the absence of any evidence that the process of screening leads to effective management of women with perinatal depression and improved mother and infant outcomes[16].

In 2007, the UK National Institute for Health and Care Excellence (NICE) produced guidelines on antenatal and postnatal mental health[2]. These set out recommendations for the detection and treatment of mental health problems during pregnancy and the postnatal period. As part of these

guidelines NICE endorsed a case-finding strategy by recommending the use of two 'ultra-brief' questions to aid the identification of perinatal depression, with the addition of a 'help' question to be asked to those women who answered 'yes' to either of the initial case-finding questions (see Box 1); these questions are often referred to as the 'Whooley' questions[17]. However, this NICE recommendation was made in the absence of any validation studies of these case-finding (Whooley) questions in a perinatal population. Instead, NICE called for a validation study to be undertaken examining the effectiveness of the Whooley questions against a diagnostic gold standard interview in women during the first postnatal year[2]. Furthermore, since the commissioning of the current study, NICE have updated their guidelines in which they continue to recommend the use of the Whooley questions during pregnancy and the postnatal period, although they have removed reference to the use of the additional help question[3].

1	"During the past month, have you often been bothered by feeling down, depressed or hopeless?"
2.	"During the past month, have you often been bothered by having little interest or pleasure in doing things?"
	A third question should be considered if the woman answers "yes" to either of the initial screening questions:
3.	"Is this something you feel you need or want help with?"

Box 1: Whooley questions for identifying perinatal depression recommended by NICE[2].

The Born and Bred in Yorkshire – PeriNatal Depression Diagnostic Accuracy study (BaBY PaNDA) therefore aims to close this evidential gap by conducting a validation study of the Whooley questions against a diagnostic gold standard interview both during pregnancy and the postnatal period. The study will also include an examination of the diagnostic validity of the EPDS as this measure is commonly used to detect postnatal depression in maternity and child services[18]. The authors have previously conducted a systematic review commissioned by the National Institute for Health Research Health Technology Assessment (NIHR HTA) of existing methods to identify postnatal depression in primary care[19]. This revealed a lack of evidence for the validity of the Whooley questions as an identification strategy for postnatal depression. This review has since been updated and found only limited evidence for the use of the Whooley questions as a case-finding strategy for postnatal depression[20].

The current study builds upon pilot work where we have tested the feasibility of longitudinal validation across the perinatal period within the UK National Health Service (NHS) maternity services. This work produced estimates of the diagnostic properties of the Whooley questions in a small but diverse sample of women during pregnancy and the early postnatal period[21]. The study found that the Whooley questions had a sensitivity of 100% (95% confidence interval [CI] 77%-100%) and a specificity of 68% (95% CI 58%-76%) during pregnancy, with similar estimates during the early postnatal period (first three postnatal months). The BaBY PaNDA study addresses the need to replicate these results in a larger sample of women representing a wider geographical population spanning different NHS trusts.

If case-finding questions are to be used to aid identification of perinatal depression in routine clinical practice, then it is important that they are acceptable to those women answering the questions and to the healthcare professionals asking the questions. At the time of commissioning the current study, previous research indicated that there were limited studies examining the acceptability to women and healthcare professionals of depression case-finding questions, such as the Whooley questions and the EPDS[19, 22], although further research has since been conducted in this area[23, 24]. The current study will determine the acceptability to women and healthcare professionals of such depression case-finding questions and will assess the potential implications for the care pathway for women diagnosed with perinatal depression. This important information will be used alongside the diagnostic estimates of the case-finding questions to inform the implementation of the NICE-endorsed case-finding strategy.

The current study also aims to investigate additional and related aspects of perinatal depression, including the relationship between depression before and after birth and co-existing psychological symptoms. Policy recommendations issued by the UK National Screening Committee (NSC) recognised the need for prospective epidemiological estimates of perinatal depression and psychological co-morbidity. The natural course of perinatal depression is under-researched, with a history of focussing on postnatal depression. Studies which have reported the under-detection of perinatal depression by healthcare professionals are largely drawn from cross-sectional studies of postnatal depression. Research is needed to determine the degree to which women with antenatal depression continue to be symptomatic in the postnatal period and the proportion of women who are identified as 'new cases' in the postnatal period.

Depression is not always experienced in isolation; epidemiological research shows that depression commonly co-exists with other common mental health disorders such as general anxiety and somatoform complaints. Assessments of depression need to recognise and assess for co-existing psychological symptoms to avoid the risk of delivering suboptimal treatment strategies. In line with this, treatment strategies, such as psychosocial interventions, need to consider the full range of co-morbid psychological symptoms if they are to be effective. NICE guidance has highlighted the importance of recognising co-existing psychological co-morbidity[25]; however, the issue of psychological comorbidity is not well understood in perinatal mental health research and the current study seeks to address this knowledge gap by assessing women for a range of common mental health disorders.

The current study also seeks to address the concern that screening for perinatal depression is an inefficient way of improving the quality of healthcare for pregnant women and new mothers. The additional health benefit of implementing screening programmes may be limited by factors such as the uptake of the screening programme and the degree to which additional identified cases are well managed and respond to treatment. A major criticism of screening programmes for mental health disorders is that they identify less severe disorders and that these identified cases will remit naturally without the need for any intervention[26] . To facilitate an understanding of the clinical and economic drivers of the cost-effectiveness of routine screening for postnatal depression, a state of the art decision model has been previously developed[19, 27]. A limitation of this model, however, was the limited availability of primary research on the diagnostic utility of depression screening questions and the lack of data on the temporal stability of screening scores and the natural history of screen-positive scores across the perinatal period. The BaBY PaNDA study will provide rich data to help adapt this existing decision model for the perinatal period and will enable us to produce robust estimates of the cost-effectiveness of a routine screening and case-finding strategy for perinatal depression.

This prospective validation study will fill important evidence gaps regarding the diagnostic utility, acceptability and cost-effectiveness of depression case-finding instruments. It will inform NICE guidance and UK NSC policy, enabling the NHS to make informed decisions on the implementation of screening and case-finding strategies and to plan services on the basis of rigorous evidence.

Research objectives

The study will combine epidemiological, psychometric, qualitative and health economic methods to meet a range of clinically-important objectives:

1. *Instrument validation:* To determine the diagnostic accuracy of the Whooley depression questions and the EPDS against a diagnostic gold standard during pregnancy (around 20 weeks gestation) and the early postnatal period (around 3-4 months after birth)
2. *Longitudinal assessment*: To assess the temporal stability of positive and negative screens between pregnancy and the early postnatal period, and to ascertain whether there is an optimal time to screen for perinatal depression
3. *Assessment of comorbidity*: To investigate the co-existence of depressive symptoms alongside other common mental health problems
4. *Evaluation of acceptability*: To determine the acceptability of the Whooley depression questions and the EPDS to expectant and new mothers and to healthcare professionals, and the potential implications for the care pathway, during the perinatal period
5. *Estimates of cost-effectiveness*: To assess the cost-effectiveness of the Whooley depression questions and the EPDS for routine screening for perinatal depression in maternity services

METHODS AND ANALYSIS

Study design

The BaBY PaNDA study is a prospective diagnostic accuracy study and is embedded within the existing Born and Bred in Yorkshire (BaBY) pregnancy and birth cohort study. The BaBY cohort recruits women during pregnancy, along with their partners and babies. Data are collected on maternal and infant health during pregnancy, labour and the neonatal period. Information on the psychological wellbeing of women and their partners is also obtained during pregnancy and the first postnatal year. The BaBY cohort study has a target population of around 13,500 births per year, with an estimated recruitment rate of >60% of women booked for delivery at each of four hospital sites (York, Hull, Harrogate and Scunthorpe & Goole).

The BaBY PaNDA study will determine the diagnostic accuracy of two depression case-finding instruments (the index tests) – the Whooley questions and the EPDS – against a validated diagnostic gold standard clinical assessment of depression, the Client Interview Schedule – Revised (CIS-R; the reference standard)[28] at two stages – once during pregnancy (around 20 weeks gestation) and once during the early postnatal period (around 3-4 months after birth). A 12 month follow-up will

also be conducted. Concurrent qualitative and cost-effectiveness evaluations will also be undertaken. The study will take place between April 2013 and June 2016.

Recruitment

Participants will be recruited through the BaBY cohort during a 14 month time period.

Inclusion criteria

Limited inclusion criteria will be applied to ensure a representative sample of pregnant women are recruited to the study. Women will be invited to take part in the study if they have consented to take part in the wider BaBY cohort and have consented to be contacted again as part of that consent; are less than 20 weeks pregnant; are aged 16 years or over; and currently live in an area covered by one of the four hospital research sites.

Exclusion criteria

Women will be excluded only if they are non-English speaking. Women with literacy difficulties will not be excluded; in such cases, all study information and questionnaires will be read out to them. Women who are over 24 weeks gestation at the time of receipt of a completed consent form will not be eligible to participate in the study.

Recruitment procedure

Recruitment will take place over a 14 month consecutive period across each of the four hospital research sites recruiting to the wider BaBY cohort: York (study coordinating site), Hull, Harrogate and Scunthorpe & Goole. All women who consent to participate in the BaBY cohort and who meet all the BaBY PaNDA inclusion criteria will be invited to take part in the study. Eligible women will be sent an information pack at round 15-18 weeks gestation; this will include an invitation letter, a summary information sheet describing the key aspects of the study, a participant information leaflet describing the study in detail, a consent form and a pre-paid return envelope. Contact details for the project team will be provided on the information leaflets, should women wish to request further information about the study. Women who wish to take part in the BaBY PaNDA study will be required to complete the consent form and return this to the research team. Women will be contacted by a member of the research team upon receipt of a completed consent form to arrange the 20 week assessment. Women who do not return a completed consent form within 2 weeks of receiving the information pack may be contacted by the research team to discuss the study and to provide them with an opportunity to ask further questions about the study.

Information about the BaBY PaNDA study (and the BaBY cohort) will be sent to all GP practices in the recruiting regions and will be displayed in locations where pregnant women attend as part of their maternity care pathway (e.g. antenatal clinics, GP surgeries).

Index tests and reference standard

The study involves validating two separate index tests against the same reference standard. The index texts and reference standard will be administered within the same session, with the index tests administered before the reference standard. For cases where it is not possible to administer the index tests and the reference standard in the same session, the reference standard will be administered within two weeks of participants completing the index tests.

Index tests

Whooley questions:

1. "During the past month, have you often been bothered by feeling down, depressed or hopeless?" ('Yes' / 'No')
2. "During the past month, have you often been bothered by having little interest or pleasure in doing things?" ('Yes' / 'No')

A 'yes' response to either of these questions will be considered a positive screen for perinatal depression and will require a response to the 'help' question:

"Is this something you feel you need or want help with?" ('Yes' / 'Yes, but not today' / 'No').

The Whooley questions have been previously validated in primary care populations[17, 29] and other clinical populations[30-32]. Since the design of the BaBY PaNDA study, they have also been validated in small perinatal populations, with sensitivity and specificity estimates in the range of 46%-100% and 65%-92%, respectively[21, 24]. The Whooley questions were selected as the primary index test as these questions are recommended by NICE to aid identification of depression during the perinatal period[2] and validation studies for these questions are limited in a perinatal population.

Edinburgh Postnatal Depression Scale (EPDS)

The EPDS[14] is a 10-item self-report questionnaire measuring depressive symptoms over the past seven days (e.g. 'I have been so unhappy that I have had difficulty sleeping', 'I have felt sad or

miserable'). Each item is scored on a four-point Likert scale (0-3), with a total score ranging from 0-30. The EPDS has a reported sensitivity of 91% and specificity of 91% when using a cut-off score of ≥13 to detect major depression in the postnatal period[33]. The EPDS was chosen as one of the index tests as it is a commonly used measure to detect postnatal depression in maternity and child services[18]. It has also been validated for use in pregnancy[34].

Reference standard

Diagnostic gold standard

The Clinical Interview Schedule – Revised (CIS-R)[28] is a self-report, computer-based interview which assesses depression severity and diagnosis, and other common mental health disorders, such as anxiety disorders, according to the International Classification of Disease (ICD-10) criteria[35].It has been validated in primary care samples with good reliability and has been used in national psychiatric morbidity surveys[28, 36]. It has also been validated for use over the telephone[37]. The CIS-R was chosen as the diagnostic gold standard due to its self-report format.

Blinding

The index tests and reference standard will be administered in the same session by one researcher. The level of potential bias is considered minimal as the EPDS (index test) and CIS-R (reference standard) are both self-report measures completed on paper (EPDS) or on a computer (CIS-R) with only minimal interaction with the researcher. To capture any potential sources of bias, an 'participant assessment record sheet' will be completed by researchers following all sessions with participants. This will include details of any questions raised by the participant during completion of the index tests and reference standard (and any other outcome measures completed as part of the session) and any information provided by the participant about their circumstances (past or current).

Blinding of outcome results will be maintained across the prenatal and postnatal stages (stages 1 and 2) with different researchers conducting these sessions for each participant, except in those instances where it may be more sensitive for the same researcher to conduct subsequent sessions.

Outcome measures and data collection

Data collection will occur at three time points during the study:

Stage 1: Prenatal (20 weeks gestation)

Stage 2: Postnatal (3-4 month post-birth)

Stage 3: Follow-up (12 months post-birth)

The main outcome measures will be the two depression case-finding instruments (as the index tests) - the Whooley questions and the EPDS. These instruments will be validated against a diagnostic gold standard – the CIS-R (as the reference standard). These three measures will be administered at stages 1, 2 and 3.

Secondary outcome measures will assess psychological comorbidity with a range of self-report questionnaires administered at stages 1, 2 and 3. These will assess symptoms of depression (Patient Health Questionnaire; PHQ-9[38]); anxiety (GAD-7[39]) and somatic symptom severity (PHQ-15[40]). Health-related quality of life and health-state utility will be assessed via the SF-12[41] and EQ5D[42]. Resource utilisation will be captured using a bespoke questionnaire completed at each of the three stages. Acceptability of the depression screening instruments will be assessed with a self-report survey originally designed to assess acceptability of the EPDS[43], later adapted to include an assessment of the Whooley questions[21], and further adapted for use in the BaBY PaNDA study. The acceptability survey will be administered at stages 1 and 2 only. Minimal biographic and demographic information will also be obtained at stage 1 only.

Outcome measures will be obtained during face-to-face interviews at stages 1 and 2. At stage 3, and for those women unable to attend a face-to-face interview at stage 2, data will be collected by telephone or a combination of telephone (diagnostic gold standard) and post (self-report questionnaires). Face-to-face interviews will be arranged for those women who specifically request this method of data collection at stage 3. Face-to-face interviews will be conducted at a time and place of the women's choosing (e.g. antenatal clinic, the women's home).

Sample size

We based the sample size calculation on a previously developed method for diagnostic accuracy studies[44]. For an expected sensitivity of 95% and a minimal acceptable lower 95% confidence interval (CI) of 80% with 0.95 probability, a total number of 50 cases is required. The estimated prevalence of perinatal depression (prenatal and postnatal) is 20%. Attrition between the prenatal and postnatal stages was estimated at 34%, based on a previous validation study of the Whooley questions in a perinatal population[20]. Therefore the sample size needed will be 379 women.

Qualitative Interviews

We will conduct a concurrent mixed-methods qualitative evaluation to determine the acceptability of the depression case-finding instruments (Whooley questions and the EPDS) to women (both

during pregnancy and the first postnatal year) and to healthcare professionals. The interviews will also explore the extent to which they capture appropriate information for effective screening of perinatal depression in routine perinatal care and the potential implications for the care pathway of delivering the depression case-finding instruments in routine care. Interviews will be conducted by a qualitative researcher.

Participant interviews

Data collection will include both a quantitative survey (acceptability survey) to be completed by all women in the study at stages 1 and 2, and in-depth semi-structured interviews to be completed with a purposive sub-sample of 25-30 women. The interview sampling framework will aim for maximum variation on the basis of socio-demographic background, age, parity, positive/negative screens on the Whooley questions and hospital research site. Women will participate in a maximum of three in-depth interviews following completion of the BaBY PaNDA outcome measures at stages 1, 2 and 3 to discuss their views of the depression case-finding instruments and, where appropriate, their experience of the care pathway. Interviews will be guided by the use of a semi-structured topic guide based on cognitive interviewing methodology[45] and open-ended probes. Women will provide their consent to be approached to take part in in-depth interviews at the point of consenting to the BaBY PaNDA study. Women who agree to participate in in-depth interviews will complete a consent form for this aspect of the study.

Health professional interviews

In-depth semi-structured single interviews will be conducted with a purposive sample of six midwives and six health visitors, to include diversity in age, professional grade, experience and hospital site. Interviews will explore health professionals' views and experience of delivering the depression case-finding instruments in routine clinical practice and their associated training needs, against descriptions of recommended routine practice and policy from health professionals in the respective hospital research site. Health professionals will be provided with an information sheet about the interviews and will be required to complete a consent form prior to conducting the interview.

DATA ANALYSIS

Statistical analysis of diagnostic accuracy data

Two-by-two contingency tables will be used to calculate sensitivity, specificity and predictive values and associated 95% confidence intervals for the Whooley questions and the EPDS against the

diagnostic gold standard (CIS-R) at stage 1 (20 weeks gestation) and stage 2 (3-4 month postnatal). Receiver operating characteristics (ROC) curves will be constructed to determine performance characteristics for the Whooley questions and the EPDS at each time point. Indeterminate and/or missing results will be summarised with respect to numbers of women and reasons (if known). The baseline characteristics of women with complete data will be compared to those of women with indeterminate and/or missing data using descriptive statistics. Predictors of non-response will be identified using a logistic regression model if there are sufficient numbers.

Based on the predictive values of the Whooley questions and the EPDS, we will identify optimum times for screening using these instruments. The temporal stability of participant responses to the Whooley questions and the EPDS between stages 1 and 2 will be explored using McNemars test. The co-existence of depressive symptoms alongside other common mental health problems at stages 1 and 2 will be summarised descriptively (mean, standard deviation, medium, minimum and maximum, and frequency and percentages at established cut points). Full details will be provided in the statistical analysis plan.

Qualitative analysis

Interviews will be audio-recorded (with participants' consent) and transcribed verbatim. Transcripts will be anonymised to ensure confidentiality. Quantitative data from the acceptability survey will be scored to produce frequency descriptive data on issues relating to acceptability and user-preference. Analysis of the qualitative data from the acceptability survey will be subjected to thematic content analysis to include coding of data using constant comparison techniques within the broader context of the existing literature.

The in-depth interviews will be examined holistically using phenomenological research methods on a case-by-case basis to describe women's and health professionals' experience in relation to their own situation and over time[46-48]. Potential sources of response error for the Whooley questions and the EPDS will be assessed using the cognitive interview approach. The interview data will also be used to further examine the findings from the acceptability survey. The health records of those women participating in in-depth interviews with a positive screen on the Whooley questions at stages 1 or 2 may be examined to triangulate their experience of the depression case-finding instruments and their care pathway.

Economic analysis

The economic evaluation will be conducted from the NHS and personal social services perspective and will include individual-level quality of life data based on the EQ-5D measure and cost data based on a bespoke resource-use questionnaire. Data recorded on the time taken to fully administer the Whooley questions and the EPDS will also be included. A hypothetical population of pregnant women managed in primary care will be evaluated using a decision analytic model consisting of two parts: (1) an identification model which reflects the diagnostic performance and administration costs of the Whooley questions and the EPDS as perinatal depression identification strategies; and (2) a treatment model which evaluates the health-related costs and outcomes (expressed as quality adjusted life years; QALYs) that may occur following administration of the depression case-finding instruments. The decision analytic model will be evaluated for true positive, false negative, true negative and false positive diagnosis groups. Using the diagnostic performance characteristics (sensitivity and specificity values) of the two depression case-finding questionnaires, the impact of true and false identification of perinatal depression and subsequent treatment of perinatal depression on costs and QALYs will be evaluated over the period of the study.

Probabilistic sensitivity analysis using the Monte Carlo simulation method[49, 50] will be undertaken to evaluate uncertainty in parameter estimates in the decision analytic model. To evaluate decision uncertainty, the simulation method will propagate uncertainty in input parameters through the model. Cost-effectiveness plane will be used to present the joint distribution of incremental costs and QALYs. Cost-effectiveness acceptability curves (CEAC) will represent the probability that the Whooley questions are cost-effective compared to the EPDS as a depression case-finding instrument for a range of willingness to pay thresholds that a UK decision-maker may consider [51].

STUDY STATUS

Recruitment of participants is completed. The first participant was enrolled in August 2013. The last participant will complete follow-up (stage 3) in January 2016.

ETHICS AND DISSEMINATION

Ethical issues

Ethical and safety considerations

As this study does not involve providing any form of intervention to the participants, we do not anticipate any major ethical concerns and consider this study low risk for participants. However, we acknowledge that some women may be vulnerable during pregnancy and the postnatal period and

may feel anxious about the identification of risk of depressive symptoms. There may also be ethical issues relating to the identification of possible cases of self-harm and/or suicide. Such issues may arise following completion of study outcomes and/or participation in qualitative interviews. Members of the research team have experience of conducting mental health studies and are well placed to deal with such ethical issues. Further, clinical members of the research team will be available to discuss any issues or concerns with researchers and/or the participant, if felt appropriate or requested. We will follow good clinical practice in monitoring risk for self-harm/suicide during researcher encounters with all participants. Robust protocols will be in place to deal with cases where risk of depression, self-harm or suicide is identified or expressed; this may involve contacting the participant's GP where necessary, with the participant's consent.

Anticipated risks and benefits

This study is considered low risk for participants. Participants will continue to receive their usual standard of maternity care, and participation in this study will not affect the standard of care they receive from their GP, midwife or health visitor. No treatment will be withheld from participants by their taking part in the study. Information about known risks and possible benefits of taking part in the study will be provided in the participant information sheet. Participants will be informed if new information comes to light which may affect their willingness to participate in the study. The participant information sheet advises potential participants that they may wish to discuss participation in the study with their GP.

Obtaining informed consent

Participants will receive an information pack about the study by post. This will contain an invitation letter, a summary information leaflet, a detailed participant information sheet and a consent form. The participant information sheet will provide contact details of the research team should participants wish to request further information about the study or ask any questions before providing their written consent. Researchers will discuss the study with participants and answer any questions during first contact with the participant following receipt of written informed consent.

Retention of study documentation

Study data will be stored in accordance with the Department of Health Sciences Data Security Policy at the University of York. Paper records will be stored in secure facilities, and all electronic records will be stored on a password protected server within the Department of Health Sciences at the University of York. Personal identifiable paper records will be stored in a separate location from

anonymised data paper records. All personal information will be destroyed at the end of the study. Anonymised data will be stored for a minimum of 20 years after the final study analysis.

Dissemination plan

We will publish the findings of this study to include (as a minimum) the diagnostic performance of the Whooley depression questions and the EPDS during pregnancy and the early postnatal period, as well as the findings from the qualitative interviews with participants and health professionals and results of the cost-effectiveness analysis. Findings will be published in peer-reviewed journals and professional journals to ensure accessibility to health researchers and clinicians. Study findings will be published using the Standards for Reporting Diagnostic Accuracy studies (STARD) guidelines[52, 53]. We will present our findings at national conferences on perinatal depression, enabling the effective dissemination of our results to a wide target audience, to include midwives, health visitors, GP and mental health professionals. We will also issue a press release to ensure coverage of our findings in the wider media. We will produce a short summary of the results for dissemination to all study participants as well as other relevant patient and other interest groups.

Contributors: EL, SA, PA, LD, CH, AK, RM, DMc and SG were responsible for the conception and design of the study and initial protocol. DM contributed to the development of the initial protocol. EL, LD, SGas, KS and BW contributed to refinements to the protocol and data acquisition. EL drafted the first draft of the manuscript. All authors contributed to and approved the final version of the manuscript.

Funding statement: This study was funded by the National Institute for Health Research Health Services and Delivery Research (NIHR HS&DR) programme (project number 11/2004/23). This paper presents independent research and the views and opinions expressed are those of the authors and do not necessarily reflect those of the HS&DR, NIHR, NHS or the Department of Health.

Competing Interests: None declared.

Patient consent: Obtained.

Ethics approval: North East – York Research Ethics Committee (Reference: 11/NE/0022).

REFERENCES

1. Murray C, Lopez A. *The global burden of disease: a comprehensive assessment of mortality and disability from disease, injuries andn risk factors in 1990*. Boston Mass: Harvard School of Public Health on behalf of the World Bank, 1996.
2. National Institute for Health and Clinical Excellence. *Antenatal and Postnatal mental health: Clinical management and service guidance. NICE clinical guideline 45.* London: The British Psychological Society & The Royal College of Psychiatrists, 2007.
3. National Institute for Health and Care Excellence. *Antenatal and postnatal mental health: Clinical management and service guidance. NICE clinical guideline 192.* London: The British Psychological Society & The Royal College of Psychiatrists, 2014.
4. Bennett HA, Einarson A, Taddio A, et al. Prevalence of depression during pregnancy: systematic review. *Obstet Gynecol* 2004;**103**(4):698-709
5. Marcus SM. Depression during pregnancy: rates, risks and consequences. *The Canadian journal of clinical pharmacology* 2009;**16**(1):e15-22
6. Gavin NI, Gaynes BN, Lohr KN, et al. Perinatal depression: a systmatic review of prevalence and incidence. *Obstet Gynecol* 2005;**106**:1071-83.
7. Boath E, Pryce A, Cox J. Postnatal depression: the impact on the family. *Journal of Reproductive and infant psychology* 1998;**16**(2-3):199-203
8. Murray L, Fiori-Cowley A, Hooper R, et al. The impact of postnatal depression and associated adversity on early mother-infant interactions and later infant outcome. *Child Dev* 1996;**67**(5):2512-26
9. Lovestone S, Kumar R. Postnatal psychiatric illness: the impact on partners. *Br J Psychiatry* 1993;**163**(2):210-16
10. Murray L, Sinclair D, Cooper P, et al. The socioemotional development of 5-year-old children of postnatally depressed mothers. *J Child Psychol Psychiatry* 1999;**40**(08):1259-71
11. Cogill S, Caplan H, Alexandra H, et al. Impact of maternal postnatal depression on cognitive development of young children. *Br Med J (Clin Res Ed)* 1986;**292**(6529):1165
12. Hearn G, Iliff A, Jones I, et al. Postnatal depression in the community. *Br J Gen Pract* 1998;**48**(428):1064
13. Secretary of State for Health. *National Service Framework - Mental Health*. London: HMSO, 1999.
14. Cox J, Holden J, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987;**150**(6):782-86
15. Shakespeare J. *Evaluation of screening for postnatal depression against the NSC handbook criteria*. London (UK): National Screening Committee, 2001
16. Henshaw C, Elliott S. *Screening for perinatal depression*: London: Jessica Kingsley Publishers, 2005.
17. 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**(7):439-45
18. Hewitt CE, Gilbody SM, Mann R, et al. Instruments to identify post-natal depression: Which methods have been the most extensively validated, in what setting and in which language? *Int J Psychiatry Clin Pract* 2010;**14**(1):72-76
19. Hewitt CE, 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**(36)
20. Mann R, Gilbody S. Validity of two case finding questions to detect postnatal depression: A review of diagnostic test accuracy. *J Affect Disord* 2011;**133**(3):388-97
21. Mann R, Adamson J, Gilbody SM. Diagnostic accuracy of case-finding questions to identify perinatal depression. *CMAJ* 2012;**184**(8):E424-E30

22. Brealey SD, Hewitt C, Green JM, et al. Screening for postnatal depression–is it acceptable to women and healthcare professionals? A systematic review and meta-synthesis. *Journal of Reproductive and infant psychology* 2010;**28**(4):328-44
23. Mann R, Adamson J, Gilbody S. The acceptability of case-finding questions to identify perinatal depression. *British Journal of Midwifery* 2015;**23**(9)
24. Darwin Z, McGowan L, Edozien LC. Identification of women at risk of depression in pregnancy: using women's accounts to understand the poor specificity of the Whooley and Arroll case finding questions in clinical practice. *Arch Womens Ment Health* 2015 doi: 10.1007/s00737-015-0508-1.
25. National Institute for Health and Clinical Excellence. *Common mental health disorders: identification and pathways to care. NICE clinical guideline 123*. London: National Institute for Health and Clinical Excellence, 2011.
26. Gilbody S, Sheldon T, Wessely S. Health policy: Should we screen for depression? *BMJ* 2006;**332**(7548):1027
27. Paulden M, Palmer S, Hewitt C, et al. Screening for postnatal depression in primary care: cost effectiveness analysis. *BMJ* 2009;**339**
28. 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**(02):465-86
29. Arroll B, Goodyear-Smith F, Kerse N, et al. Effect of the addition of a “help” question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *BMJ* 2005;**331**(7521):884
30. Mohr D, Hart S, Julian L, et al. Screening for depression among patients with multiple sclerosis: two questions may be enough. *Mult Scler* 2007;**13**(2):215-19
31. McManus D, Pipkin SS, Whooley MA. Screening for depression in patients with coronary heart disease (data from the Heart and Soul Study). *Am J Cardiol* 2005;**96**(8):1076
32. Mallen CD, Peat G. Screening older people with musculoskeletal pain for depressive symptoms in primary care. *Br J Gen Pract* 2008;**58**(555):688
33. Gaynes BN, Gavin N, Meltzer-Brody S, et al. Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes: *Evidence Report/Technology Assessment (Summary)*, 2005;1-8
34. Murray D, Cox JL. Screening for depression during pregnancy with the Edinburgh Depression Scale (EDDS). *Journal of Reproductive and infant psychology* 1990;**8**(2):99-107
35. Gruenberg AM, Goldstein RD, Pincus HA. *Classification of depression: research and diagnostic criteria: DSM-IV and ICD-10*. Weinheim: Wiley, 2005.
36. Singleton N, Bumpstead R, O'Brien M, et al. *Psychiatric morbidity among adults living in private households, 2000*. London, UK: The Stationery Office, 2001.
37. Evans M, Kessler D, Lewis G, et al. Assessing mental health in primary care research using standardized scales: can it be carried out over the telephone? *Psychol Med* 2004;**34**(1):157-62
38. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999;**282**(18):1737-44
39. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;**166**(10):1092
40. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med* 2002;**64**(2):258-66
41. Ware Jr JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;**34**(3):220-33
42. EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**(3):199-208
43. Gemmill AW, Leigh B, Ericksen J, et al. A survey of the clinical acceptability of screening for postnatal depression in depressed and non-depressed women. *BMC Public Health* 2006;**6**(1):211
44. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005;**58**(8):859

45. Tourangeau R. Cognitive science and survey methods. In: Jabine TB, Straf ML, Tanur JM, Tourangeau R, eds. *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press 1984:73-100
46. Moustakas C. *Phenomenological research methods*. London: Sage Publications, Inc. 1994.
47. Silverman D. *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction* (Third edition). London: SAGE Publications Ltd, 2006
48. Balls P. Phenomenology in nursing research: methodology, interviewing and transcribing. *Nurs Times* 2009;**105**(32-33):30
49. Briggs A, Claxton C, Sculpher MJ. *Decision modelling for health economic evaluation*: Oxford: Oxford University Press, 2006.
50. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ* 2005;**14**(4):339-47
51. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001;**10**(8):779-87
52. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;**351:h5527** doi: http://dx.doi.org/10.1136/bmj.h5527
53. Bossuyt PM, Reitsma JB, E Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;**326:41** doi: http://dx.doi.org/10.1136/bmj.326.7379.41.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4-7
	4	Study objectives and hypotheses	8
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	8
Participants	**6**	Eligibility criteria	9
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	9
	8	Where and when potentially eligible participants were identified (setting, location and dates)	9
	9	Whether participants formed a consecutive, random or convenience series	9
Test methods	**10a**	Index test, in sufficient detail to allow replication	10-11
	10b	Reference standard, in sufficient detail to allow replication	11
	11	Rationale for choosing the reference standard (if alternatives exist)	11
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	10
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	11
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	11
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	13-14
	15	How indeterminate index test or reference standard results were handled	13-14
	16	How missing data on the index test and reference standard were handled	13-14
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	
	18	Intended sample size and how it was determined	12
RESULTS			
Participants	**19**	Flow of participants, using a diagram	N/A
	20	Baseline demographic and clinical characteristics of participants	N/A
	21a	Distribution of severity of disease in those with the target condition	N/A
	21b	Distribution of alternative diagnoses in those without the target condition	N/A
	22	Time interval and any clinical interventions between index test and reference standard	N/A
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	N/A
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	N/A
	25	Any adverse events from performing the index test or the reference standard	N/A
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	N/A
	27	Implications for practice, including the intended use and clinical role of the index test	N/A
OTHER INFORMATION			
	28	Registration number and name of registry	N/A
	29	Where the full study protocol can be accessed	
	30	Sources of funding and other support; role of funders	18

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Identification of depression in women during pregnancy and the early postnatal period using the Whooley questions and the Edinburgh Postnatal Depression Scale: protocol for the Born and Bred in Yorkshire - PeriNatal Depression Diagnostic Accuracy (BaBY PaNDA) study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011223.R1
Article Type:	Protocol
Date Submitted by the Author:	31-Mar-2016
Complete List of Authors:	Littlewood, Elizabeth; The University of York, Health Sciences Ali, Shehzad; University of York, Department of Health Sciences Ansell, Pat; University of York, Department of Health Sciences Dyson, Lisa; University of York, York Trials Unit Gascoyne, Samantha; University of York, Department of Health Sciences Hewitt, Catherine; The University of York, Health Sciences Keding, Ada; University of York, Health Sciences Mann, Rachel; University of York, Social Policy Research Unit McMillan, Dean; University of York, Department of Health Sciences Morgan, Deborah Swan, Kelly; University of York, Department of Health Sciences Waterhouse, Bev; Calderdale and Huddersfield NHS Foundation Trust, Children, Women and Families Division Gilbody, Simon; The University of York, Department of Health Sciences
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Health services research
Keywords:	diagnostic accuracy, Whooley questions, perinatal depression, Edinburgh Postnatal Depression Scale, screening

TITLE: Identification of depression in women during pregnancy and the early postnatal period using the Whooley questions and the Edinburgh Postnatal Depression Scale: protocol for the Born and Bred in Yorkshire – PeriNatal Depression Diagnostic Accuracy (BaBY PaNDA) study

Elizabeth Littlewood[1]*, Shehzad Ali[1], Pat Ansell[2], Lisa Dyson[3], Samantha Gascoyne[1], Catherine Hewitt[3], Ada Keding[3], Rachel Mann[4], Dean McMillan[1,5], Deborah Morgan[6], Kelly Swan[2], Bev Waterhouse[7], Simon Gilbody[1,5] on behalf of the BaBY PaNDA study team.

Author affiliations

[1]Mental Health and Addiction Research Group, Department of Health Sciences, University of York, York, YO10 5DD, UK

[2]Epidemiology and Cancer Statistics Research Group, Department of Health Sciences, University of York, York, YO10 5DD, UK

[3]York Trials Unit, Department of Health Sciences, University of York, York, YO10 5DD, UK

[4]Social Policy Research Unit, University of York, York, YO10 5DD, UK

[5]Hull York Medical School, University of York, York, YO10 5DD, UK

[6]PNI-UK, Ashbourne, Derbyshire, DE6 5NB, UK

[7]Children, Women & Families Division, Calderdale and Huddersfield NHS Foundation Trust, Calderdale Royal Hospital, Salterhebble, Halifax, HX3 0PW, UK

***Corresponding author**

Elizabeth Littlewood

Mental Health and Addiction Research Group

Department of Health Sciences

ARRC Building, Area 4

University of York

York

YO10 5DD, UK

Email: liz.littlewood@york.ac.uk

Tel: +44 (0)1904 321828

Key words: diagnostic accuracy; Whooley questions; perinatal depression; Edinburgh Postnatal Depression Scale, screening

Word count: 5,170; **Abstract; Tables: 1**

ABSTRACT

Introduction

Perinatal depression is well recognised as a mental health condition but less than 50% of cases are identified by healthcare professionals in routine clinical practice. The Edinburgh Postnatal Depression Scale (EPDS) is often used to detect symptoms of postnatal depression in maternity and child services. The National Institute of Health and Care Excellence (NICE) recommends two 'ultra-brief' case-finding questions (the Whooley questions) to aid identification of depression during the perinatal period, but this recommendation was made in the absence of any validation studies in a perinatal population. Limited research exists on the acceptability of these depression case-finding instruments and the cost-effectiveness of routine screening for perinatal depression.

Methods and analysis

The diagnostic accuracy of the Whooley questions and the EPDS will be determined against a reference standard (the Client Interview Schedule – Revised) during pregnancy (around 20 weeks) and the early postnatal period (around 3-4 months postpartum) in a sample of 379 women. Further outcome measures will assess a range of psychological comorbidities, health related quality of life and resource utilisation. Women will be followed up 12 months postnatally. The sensitivity, specificity and predictive values of the Whooley questions and the EPDS will be calculated against the reference standard at 20 weeks pregnancy and 3-4 months postpartum. Acceptability of the depression case-finding instruments to women and healthcare professionals will involve in-depth qualitative interviews. An existing decision analytic model will be adapted to determine the cost-effectiveness of routine screening for perinatal depression.

Ethics and dissemination

This study is considered low risk for participants. Robust protocols will deal with cases where risk of depression, self-harm or suicide is identified. The protocol received favourable ethical opinion from the North East – York Research Ethics Committee (reference: 11/NE/0022). The study findings will be published in peer-reviewed journals and presented at relevant conferences.

Strengths and limitations of this study

- This study will fill an important evidence gap regarding the diagnostic utility of depression case-finding instruments for the identification of perinatal depression.
- The study findings will be informed by qualitative interviews conducted with women and healthcare professionals regarding their acceptability of depression case-finding instruments administered during the perinatal period.
- An existing decision analytic model will be updated with current diagnostic accuracy estimates of two depression case-finding instruments, providing an up-to-date estimate of the cost-effectiveness of a perinatal depression screening strategy.
- The study findings will inform policy decisions on the implementation of screening and case-finding strategies for the identification of perinatal depression.
- The study spans four NHS trusts which may implement differing policies regarding the identification of and referral processes for perinatal depression during the perinatal period.

INTRODUCTION

Depression accounts for the greatest burden of disease of all mental health problems and is estimated to become the second largest cause of global disability by 2020[1]. It is well recognised that perinatal depression, that is depression experienced during pregnancy and/or the postnatal period (up to one year after birth), is an important category of depression in its own right, with specific guidance provided on the identification and clinical management of the condition[2, 3].

Prevalence rates of perinatal depression vary. Estimates indicate that approximately 7.4%-20% of women experience depression at some stage during pregnancy[4-6] with depression during the postnatal period affecting up to 22% of women[6]. Perinatal depression is associated with a range of adverse outcomes. Evidence suggests an association between depression experienced during pregnancy (prenatal depression) and adverse neonatal outcomes, poor self-reported health, substance abuse and alcohol abuse, and poor usage of antenatal care services[5]. Postnatal depression has been shown to have a substantial impact on the mother and her partner[7], mother-baby interactions[8], the family[9] and on the longer-term emotional and cognitive development of the baby[10], particularly when depression occurs in the first year of life[11].

Although perinatal depression is well recognised as a mental health condition, it often goes undetected; with healthcare professionals detecting less than 50% of cases in routine clinical practice[12]. The National Service Framework (NSF) states that local protocols should be in place for the management of postnatal depression[13], promoting the use of case-finding or screening strategies to aid identification of depression during the perinatal period. This has led to the routine or ad-hoc administration of self-report measures such as the Edinburgh Postnatal Depression Scale (EPDS)[14]. Screening or case-finding strategies such as those advocated by the NSF[13] have since come under scrutiny[15, 16] and have been criticised on a number of factors. Criticisms of the proposed strategies are based on the ethics of mass screening, concerns regarding the psychometric properties of available screening or case-finding instruments (such as variations in diagnostic accuracy estimates and choice of recommended cut-off points for such instruments), the acceptability of such screening or case-finding strategies to patients and healthcare professionals, the paucity of evidence for the cost-effectiveness of screening or case-finding strategies (particularly the costs associated with the management of incorrectly identified cases of perinatal depression), and the absence of any evidence that the process of screening leads to effective management of women with perinatal depression and improved mother and infant outcomes[16-18].

In 2007, the UK National Institute for Health and Care Excellence (NICE) produced guidelines on antenatal and postnatal mental health[2]. These set out recommendations for the detection and treatment of mental health problems during pregnancy and the postnatal period. As part of these guidelines NICE endorsed a case-finding strategy by recommending the use of two 'ultra-brief' questions to aid the identification of perinatal depression, with the addition of a 'help' question to be asked to those women who answered 'yes' to either of the initial case-finding questions (see Box 1); these questions are often referred to as the 'Whooley' questions[19]. However, this NICE recommendation was made in the absence of any validation studies of these case-finding (Whooley) questions in a perinatal population. Instead, NICE called for a validation study to be undertaken examining the effectiveness of the Whooley questions against a diagnostic gold standard interview in women during the first postnatal year[2]. Furthermore, since the commissioning of the current study, NICE have updated their guidelines in which they continue to recommend the use of the Whooley questions during pregnancy and the postnatal period, although they have removed reference to the use of the additional help question[3].

1 "During the past month, have you often been bothered by feeling down, depressed or hopeless?" ('Yes' / 'No')

2. "During the past month, have you often been bothered by having little interest or pleasure in doing things?" ('Yes' / 'No')

A third question should be considered if the woman answers "yes" to either of the initial screening questions:

3. "Is this something you feel you need or want help with?" ('Yes' / 'Yes, but not today' / 'No')

Box 1: Whooley questions for identifying perinatal depression recommended by NICE[2].

The Born and Bred in Yorkshire – PeriNatal Depression Diagnostic Accuracy study (BaBY PaNDA) therefore aims to close this evidential gap by conducting a validation study of the Whooley questions against a reference standard (the Client Interview Schedule – Revised; CIS-R)[20] during pregnancy and the postnatal period. Given that the EPDS is the measure most commonly used to detect symptoms of postnatal depression in maternity and child services[21], the study will also include a comparative examination of the diagnostic validity of the EPDS.. The authors have previously conducted a systematic review commissioned by the National Institute for Health Research Health Technology Assessment Programme (NIHR HTA) of existing methods to identify postnatal depression in primary care[18]. This revealed a lack of evidence for the validity of the Whooley questions as an identification strategy for postnatal depression. This review has since been

updated and found only limited evidence for the use of the Whooley questions as a case-finding strategy for postnatal depression[22].

The current study builds upon pilot work where we have tested the feasibility of longitudinal validation across the perinatal period within the UK National Health Service (NHS) maternity services. This work produced estimates of the diagnostic properties of the Whooley questions in a small but diverse sample of 152 women during pregnancy and the early postnatal period[23]. The study found that the Whooley questions had a sensitivity of 100% (95% confidence interval [CI] 77%-100%) and a specificity of 68% (95% CI 58%-76%) during pregnancy, with similar estimates during the early postnatal period (first three postnatal months). Similar positive likelihood ratios were found during pregnancy (3.03) and the early postnatal period (2.73), as was the case for the negative likelihood ratios (0.041 during pregnancy, 0.042 postnatally). The BaBY PaNDA study addresses the need to replicate these results in a larger sample of women representing a wider geographical population spanning different NHS trusts.

If case-finding questions are to be used to aid identification of perinatal depression in routine clinical practice, then it is important that they are acceptable to those women answering the questions and to the healthcare professionals asking the questions. At the time of commissioning the current study, previous research indicated that there were limited studies examining the acceptability to women and healthcare professionals of depression case-finding questions, such as the Whooley questions and the EPDS[18, 24], although further research has since been conducted in this area[25, 26]. The current validation study will therefore also include an assessment of the acceptability to women and healthcare professionals of such depression case-finding questions and will assess the potential implications for the care pathway for women diagnosed with perinatal depression. This important information will be used alongside the diagnostic estimates of the case-finding questions to inform the implementation of the NICE-endorsed case-finding strategy.

The current study also aims to investigate additional and related aspects of perinatal depression, including the relationship between depression before and after birth and co-existing psychological symptoms. Policy recommendations issued by the UK National Screening Committee (NSC) recognised the need for prospective epidemiological estimates of perinatal depression and psychological co-morbidity. Research investigating the natural course of perinatal depression is somewhat limited. Findings from a large longitudinal community sample (the Avon Longitudinal Study of Parents and Children; ALSPAC) suggest higher rates of depressive symptoms in women (as

measured by the EPDS) during pregnancy than during the postnatal period (up to eight months)[27]. Studies which have reported the under-detection of perinatal depression by healthcare professionals are largely drawn from cross-sectional studies of postnatal depression. Further research is needed to determine the degree to which women with prenatal depression continue to be symptomatic in the postnatal period and the proportion of women who are identified as 'new cases' in the postnatal period.

Depression is not always experienced in isolation; epidemiological research shows that depression commonly co-exists with other common mental health disorders such as general anxiety and somatoform complaints. Assessments of depression need to recognise and assess for co-existing psychological symptoms to avoid the risk of delivering suboptimal treatment strategies. In line with this, treatment strategies, such as psychosocial interventions, need to consider the full range of co-morbid psychological symptoms if they are to be effective. NICE guidance has highlighted the importance of recognising co-existing psychological co-morbidity[28]; however, the issue of psychological comorbidity is not well understood in perinatal mental health research and the current study seeks to address this knowledge gap by assessing women for a range of common mental health disorders.

The current study also seeks to address the concern that screening for perinatal depression is an inefficient way of improving the quality of healthcare for pregnant women and new mothers. The additional health benefit of implementing screening programmes may be limited by factors such as the uptake of the screening programme and the degree to which additional identified cases are well managed and respond to treatment. A major criticism of screening programmes for mental health disorders is that they identify less severe disorders and that these identified cases will remit naturally without the need for any intervention[29]. To facilitate an understanding of the clinical and economic drivers of the cost-effectiveness of routine screening for postnatal depression, a decision model has been previously developed[18, 30]. A limitation of this model, however, was the limited availability of primary research on the diagnostic utility of depression screening questions and the lack of data on the temporal stability of screening scores and the natural history of screen-positive scores across the perinatal period. The BaBY PaNDA study will provide rich data to help adapt and update this existing decision model for the perinatal period and will enable us to produce robust real-world estimates of the cost-effectiveness of a routine screening and case-finding strategy for perinatal depression.

This prospective validation study will fill important evidence gaps regarding the diagnostic utility, acceptability and cost-effectiveness of depression case-finding instruments. It will inform NICE guidance and UK NSC policy, enabling the NHS to make informed decisions on the implementation of screening and case-finding strategies and to plan services on the basis of rigorous evidence.

Research objectives

The study will combine epidemiological, psychometric, qualitative and health economic methods to meet a range of clinically-important objectives:

1. *Instrument validation:* To determine the diagnostic accuracy of the Whooley depression questions and the EPDS against a reference standard during pregnancy (around 20 weeks gestation) and the early postnatal period (around 3-4 months after birth)
2. *Longitudinal assessment*: To assess the temporal stability of positive and negative screens between pregnancy and the early postnatal period, and to ascertain whether there is an optimal time to screen for perinatal depression
3. *Assessment of comorbidity*: To investigate the co-existence of depressive symptoms alongside other common mental health problems
4. *Evaluation of acceptability*: To determine the acceptability of the Whooley depression questions and the EPDS to expectant and new mothers and to healthcare professionals, and the potential implications for the care pathway, during the perinatal period
5. *Estimates of cost-effectiveness*: To assess the cost-effectiveness of the Whooley depression questions and the EPDS for routine screening for perinatal depression in maternity services

METHODS AND ANALYSIS

Study design

The BaBY PaNDA study is a prospective diagnostic accuracy study and is embedded within the existing Born and Bred in Yorkshire (BaBY) pregnancy and birth cohort study. The BaBY cohort recruits women during pregnancy, along with their partners and babies. Data are collected on maternal and infant health during pregnancy, labour and the neonatal period. Information on the psychological wellbeing of women and their partners is also obtained during pregnancy and the first postnatal year. The BaBY cohort study has a target population of around 13,500 births per year, with an estimated recruitment rate of >60% of women booked for delivery at each of four hospital sites (York, Hull, Harrogate and Scunthorpe & Goole).

The BaBY PaNDA study will determine the diagnostic accuracy of two depression case-finding instruments (the index tests) – the Whooley questions and the EPDS – against a validated assessment of depression, the CIS-R (the reference standard)[20] at two stages – once during pregnancy (around 20 weeks gestation) and once during the early postnatal period (around 3-4 months after birth). A 12 month follow-up will also be conducted. Concurrent qualitative and cost-effectiveness evaluations will also be undertaken. The study will take place between April 2013 and June 2016.

Recruitment

Women will be recruited through the wider BaBY cohort during a 14 month time period.

Inclusion criteria

Limited inclusion criteria will be applied to ensure a representative sample of pregnant women are recruited to the study. Eligible women will be identified from the population of women taking part in the wider BaBY cohort study (described above). Pregnant women will be invited to take part in the study if they have consented to take part in the wider BaBY cohort and have consented to be contacted again as part of that consent; are less than 20 weeks pregnant; are aged 16 years or over; and currently live in an area covered by one of the four hospital research sites.

Exclusion criteria

Women will be excluded only if they are non-English speaking. Women with literacy difficulties will not be excluded; in such cases, all study information and questionnaires will be read out to them by the study researchers. Women who are over 24 weeks gestation at the time of receipt of a completed consent form will not be eligible to participate in the study.

Recruitment procedure

Recruitment will take place over a 14 month consecutive period across each of the four hospital research sites recruiting to the wider BaBY cohort: York (study coordinating site), Hull, Harrogate and Scunthorpe & Goole. All women who consent to participate in the BaBY cohort and who meet all the BaBY PaNDA inclusion criteria (including having provided consent to be contacted again) will be invited to take part in the study. Eligible women will be sent an information pack at round 15-18 weeks gestation; this will include an invitation letter, a summary information sheet describing the key aspects of the study, a participant information leaflet describing the study in detail, a consent form and a pre-paid return envelope. Contact details for the project team will be provided on the

information leaflets, should women wish to request further information about the study. Women who wish to take part in the BaBY PaNDA study will be required to complete the consent form and return this to the research team. Women will be contacted by a member of the research team upon receipt of a completed consent form to arrange the 20 week assessment. Women who do not return a completed consent form within 2 weeks of receiving the information pack may be contacted by the research team to discuss the study and to provide them with an opportunity to ask further questions about the study.

Information about the BaBY PaNDA study (and the BaBY cohort) will be sent to all GP practices in the recruiting regions and will be displayed in locations where pregnant women attend as part of their maternity care pathway (e.g. antenatal clinics, GP surgeries).

Index tests and reference standard

The study involves validating two separate index tests against the same reference standard at two separate time-points: 20 weeks pregnancy and 3-4 months post-birth. The index texts and reference standard will be administered within the same session by one researcher, with the index tests administered before the reference standard. For cases where it is not possible to administer the index tests and the reference standard in the same session, the reference standard will be administered within two weeks of participants completing the index tests. The index tests and reference standard will be administered during face-to-face interviews or over the telephone and will be conducted at a time and location according to the woman's preference (e.g. antenatal clinic, the woman's home).

Index tests

Whooley questions

Women will be asked the Whooley questions (see Box 1) by a study researcher. A 'yes' response to either of questions 1 or 2 will be considered a positive screen for perinatal depression and will require a response to the 'help' question (question 3).

The Whooley questions have been previously validated in primary care populations[19, 31] and other clinical populations[32-34]. Since the design of the BaBY PaNDA study, they have also been validated in small perinatal populations, with sensitivity and specificity estimates in the range of 46%-100% and 65%-92%, respectively[23, 26]. The Whooley questions were selected as the primary index test as these questions are recommended by NICE to aid identification of depression during

the perinatal period[2] and validation studies for these questions are limited in a perinatal population.

Edinburgh Postnatal Depression Scale (EPDS)

Women will be asked to self-complete the EPDS[14]; this is a 10-item self-report questionnaire measuring depressive symptoms over the past seven days (e.g. 'I have been so unhappy that I have had difficulty sleeping', 'I have felt sad or miserable'). Each item is scored on a four-point Likert scale (0-3), with a total score ranging from 0-30. The EPDS has a reported sensitivity of 91% and specificity of 91% when using a cut-off score of ≥13 to detect major depression in the postnatal period[35]. The EPDS was chosen as one of the index tests as it is a commonly used measure to detect symptoms of postnatal depression in maternity and child services[21] and is widely used in research in perinatal mental health. It has also been validated for use in pregnancy[36].

Reference standard

Clinical Interview Schedule – Revised

Women will be asked to self-complete the computer-based version of the CIS-R [20]. The CIS-R is a fully structured assessment which assesses 14 areas of symptoms, including depression, anxiety, sleep, fatigue, panic, phobias and compulsions/obsessions, and generates diagnostic categories (including depression severity and diagnosis), according to the International Classification of Disease (ICD-10) criteria[37]. Study researchers will be trained in the use and delivery of the CIS-R.

The CIS-R has been validated in primary care samples with good reliability and has been used in national psychiatric morbidity surveys[20, 38]. It has also been validated for use over the telephone[39]. The CIS-R was chosen as the reference standard due to its self-report format.

Blinding of outcome results across index tests and reference standard

The index tests and reference standard will be administered in the same session by one researcher. Within a session, the level of potential bias is considered minimal as the EPDS (index test) and CIS-R (reference standard) are both self-report measures completed on paper (EPDS) or on a computer (CIS-R) with only minimal interaction with the researcher. To capture any potential sources of bias, a 'participant assessment record sheet' will be completed by researchers following all sessions with participants. This will include details of any questions raised by the participant during completion of the index tests and reference standard (and any other outcome measures completed as part of the session) and any information provided by the participant about their circumstances (past or current).

Blinding of outcome results of the index tests and reference standard will be maintained across the two time-points (20 weeks pregnancy and 3-4 months postpartum) with different researchers conducting these sessions for each participant, except in those instances where it may be more sensitive for the same researcher to conduct subsequent sessions.

Outcome measures and data collection

Data collection will occur at three time points during the study:

Stage 1: Prenatal (20 weeks pregnancy)

Stage 2: Postnatal (3-4 month post-partum)

Stage 3: Follow-up (12 months post-partum)

The main outcome measures will be the two depression case-finding instruments (as the index tests) - the Whooley questions and the EPDS. These instruments will be validated against the CIS-R (as the reference standard). These three measures will be administered at stages 1, 2 and 3.

Further outcome measures will assess a range of psychological comorbidities with a number of self-report questionnaires administered at stages 1, 2 and 3. These will assess symptoms of depression (Patient Health Questionnaire; PHQ-9[40]); anxiety (GAD-7[41]) and somatic symptom severity (PHQ-15[42]). The CIS-R will also be used to identify other common mental health disorders, including panic disorder, obsessive-compulsive disorder and phobias. Health-related quality of life and health-state utility will be assessed via the SF-12[43] and EQ5D[44]. Resource utilisation will be captured using a bespoke questionnaire completed at each of the three stages. Acceptability of the depression case-finding instruments to women will be assessed with a self-report acceptability survey originally designed to assess acceptability of the EPDS[45], later adapted to include an assessment of the Whooley questions[23], and further adapted for use in the BaBY PaNDA study. The acceptability survey will be administered at stages 1 and 2 only. Acceptability of the depression case-finding instruments (to both women and healthcare professionals) will also be determined via in-depth qualitative interviews (see qualitative interviews section for further detail on the assessment of acceptability). Minimal biographic and demographic information will also be obtained at stage 1 only.

Outcome measures will be obtained during face-to-face interviews at stages 1 and 2. At stage 3, and for those women unable to attend a face-to-face interview at stage 2, data will be collected by telephone or a combination of telephone (CIS-R) and post (self-report questionnaires). Face-to-face interviews will be arranged for those women who specifically request this method of data collection at stage 3. Face-to-face interviews will be conducted at a time and place of the women's choosing (e.g. antenatal clinic, the women's home). Women are advised via the participant information leaflet and during initial discussions with the study researchers that each session will last approximately 30-40 minutes.

Sample size

We based the sample size calculation on a previously developed method for diagnostic accuracy studies[46]. For an expected sensitivity of 95% and a minimal acceptable lower 95% confidence interval (CI) of 80% with 0.95 probability, a total number of 50 cases of women with depression in the perinatal period is required. The estimated prevalence of perinatal depression (prenatal and postnatal) is 20%[6]. Attrition between the prenatal and postnatal stages was estimated at 34%, based on a previous validation study of the Whooley questions in a perinatal population[22]. Therefore the sample size needed will be 379 women.

Qualitative Interviews

We will conduct a concurrent mixed-methods qualitative evaluation to determine the acceptability of the Whooley questions and the EPDS to women (both during pregnancy and the first postnatal year) and to healthcare professionals. The interviews will also explore the extent to which they capture appropriate information for effective screening of perinatal depression in routine perinatal care and the potential implications for the care pathway of delivering the depression case-finding instruments in routine care. Interviews will be conducted by a qualitative researcher.

Participant interviews

Data collection will include both a quantitative survey (the adapted acceptability survey) to be completed by all women in the study at stages 1 and 2, and in-depth semi-structured interviews to be completed with a purposive sub-sample of 25-30 women. The interview sampling framework will aim for maximum variation on the basis of socio-demographic background, age, parity, positive/negative screens on the Whooley questions and hospital research site. Women will participate in a maximum of three in-depth interviews following completion of the BaBY PaNDA outcome measures at each of stages 1, 2 and 3 to discuss their views of the depression case-finding

instruments and their associated experience of the care pathway. Interviews will be conducted on a subsequent and separate occasion to completion of the BaBY PaNDA outcome measures and will be conducted at a time and location according to the woman's preference. Interviews will be guided by the use of a semi-structured topic guide based on cognitive interviewing methodology[47] and open-ended probes. Women will provide their consent to be approached to take part in in-depth interviews at the point of consenting to the BaBY PaNDA study. Women who agree to participate in in-depth interviews will complete a consent form for this aspect of the study.

Health professional interviews

In-depth semi-structured single interviews will be conducted with a purposive sample of six midwives and six health visitors, to include diversity in age, professional grade, experience and hospital site. Interviews will explore health professionals' views and experience of using the depression case-finding instruments as part of routine clinical practice within their NHS trust and their associated training needs, against descriptions of recommended routine practice and policy from health professionals in the respective hospital research site. Health professionals will be provided with an information sheet about the interviews and will be required to complete a consent form prior to conducting the interview.

DATA ANALYSIS

Statistical analysis of diagnostic accuracy data

Two-by-two contingency tables will be used to calculate sensitivity, specificity and predictive values and associated 95% confidence intervals for the Whooley questions and the EPDS against the reference standard (CIS-R) at stage 1 (20 weeks pregnancy) and stage 2 (3-4 months postnatal). Receiver operating characteristics (ROC) curves will be constructed to determine performance characteristics for the Whooley questions and the EPDS at each time point. Indeterminate and/or missing results will be summarised with respect to numbers of women and reasons (if known). The baseline characteristics of women with complete data will be compared to those of women with indeterminate and/or missing data using descriptive statistics. Predictors of non-response will be identified using a logistic regression model if there are sufficient numbers.

Based on the predictive values of the Whooley questions and the EPDS, we will establish which of the two time-points (20 weeks pregnancy or 3-4 months postnatal) is better to establish perinatal mental health. The temporal stability of participant responses to the Whooley questions and the EPDS between stages 1 and 2 will be explored using McNemars test. The co-existence of depressive

symptoms alongside other common mental health problems at stages 1 and 2 will be summarised descriptively (mean, standard deviation, medium, minimum and maximum, and frequency and percentages at established cut points). Full details will be provided in the statistical analysis plan.

Qualitative analysis

Interviews will be audio-recorded (with participants' consent) and transcribed verbatim. Transcripts will be anonymised to ensure confidentiality. Quantitative data from the acceptability survey will be scored to produce frequency descriptive data on issues relating to acceptability and user-preference. Analysis of the qualitative data from the acceptability survey will be subjected to thematic content analysis to include coding of data using constant comparison techniques within the broader context of the existing literature.

The in-depth interviews will be examined holistically using phenomenological research methods on a case-by-case basis to describe women's and health professionals' experience in relation to their own situation and over time[48-50]. Potential sources of response error for the Whooley questions and the EPDS will be assessed using the cognitive interview approach. The interview data will also be used to further examine the findings from the acceptability survey. The health records of those women participating in in-depth interviews with a positive screen on the Whooley questions at stages 1 or 2 may be examined to triangulate their experience of the depression case-finding instruments and their care pathway.

Economic analysis

The economic evaluation will be conducted from the NHS and personal social services perspective and will include individual-level quality of life data based on the EQ-5D measure and cost data based on a bespoke resource-use questionnaire. Data recorded on the time taken to fully administer the Whooley questions and the EPDS will also be included. A hypothetical population of pregnant women managed in primary care will be evaluated using a decision analytic model consisting of two parts: (1) an identification model which reflects the diagnostic performance and administration costs of the Whooley questions and the EPDS as perinatal depression identification strategies; and (2) a treatment model which evaluates the health-related costs and outcomes (expressed as quality adjusted life years; QALYs) that may occur following administration of the depression case-finding instruments. The decision analytic model will be evaluated for true positive, false negative, true negative and false positive diagnosis groups. Using the diagnostic performance characteristics (sensitivity and specificity values) of the two depression case-finding questionnaires, the impact of

true and false identification of perinatal depression and subsequent treatment of perinatal depression on costs and QALYs will be evaluated over the period of the study.

Probabilistic sensitivity analysis using the Monte Carlo simulation method[51, 52] will be undertaken to evaluate uncertainty in parameter estimates in the decision analytic model. To evaluate decision uncertainty, the simulation method will propagate uncertainty in input parameters through the model. Cost-effectiveness plane will be used to present the joint distribution of incremental costs and QALYs. Cost-effectiveness acceptability curves (CEAC) will represent the probability that the Whooley questions are cost-effective compared to the EPDS as a depression case-finding instrument for a range of willingness to pay thresholds that a UK decision-maker may consider[53].

The decision model will be developed with reference to NICE guidelines for antenatal and postnatal mental health[2, 3] to reflect recommended clinical practice and to ensure that the decision model is realistic and relevant to clinical context.

STUDY STATUS

Recruitment of participants is completed. The first participant was enrolled in August 2013. The last participant will complete follow-up (stage 3) in January 2016.

ETHICS AND DISSEMINATION

Ethical issues

Ethical and safety considerations

As this study does not involve providing any form of intervention, we do not anticipate any major ethical concerns and consider this study low risk for participants. However, we acknowledge that some women may be vulnerable during pregnancy and the postnatal period and may feel anxious about the identification of risk of depressive symptoms. There may also be ethical issues relating to the identification of possible cases of self-harm and/or suicide. Such issues may arise following completion of study outcomes and/or participation in qualitative interviews. Members of the research team have experience of conducting mental health studies and are well placed to deal with such ethical issues. Further, clinical members of the research team will be available to discuss any issues or concerns with researchers and/or the participant, if felt appropriate or requested. We will follow good clinical practice in monitoring risk for self-harm/suicide during researcher encounters with all participants. Robust protocols will be in place to deal with cases where risk of depression, self-harm or suicide is identified or expressed; this may involve contacting the participant's GP where necessary, with the participant's consent.

Anticipated risks and benefits

This study is considered low risk for participants. Participants will continue to receive their usual standard of maternity care, and participation in this study will not affect the standard of care they receive from their GP, midwife or health visitor. No treatment will be withheld from participants by their taking part in the study. Information about known risks and possible benefits of taking part in the study will be provided in the participant information sheet. Participants will be informed if new information comes to light which may affect their willingness to participate in the study. The participant information sheet advises potential participants that they may wish to discuss participation in the study with their GP.

Obtaining informed consent

Eligible participants will receive an information pack about the study by post. This will contain an invitation letter, a summary information leaflet, a detailed participant information sheet and a consent form. The participant information sheet will provide contact details of the research team should participants wish to request further information about the study or ask any questions before providing their written consent. Researchers will discuss the study with participants and answer any questions during first contact with the participant following receipt of written informed consent.

Retention of study documentation

Study data will be stored in accordance with the Department of Health Sciences Data Security Policy at the University of York. Paper records will be stored in secure facilities, and all electronic records will be stored on a password protected server within the Department of Health Sciences at the University of York. Personal identifiable paper records will be stored in a separate location from anonymised data paper records. All personal information will be destroyed at the end of the study. Anonymised data will be stored for a minimum of 20 years after the final study analysis.

Dissemination plan

We will publish the findings of this study to include (as a minimum) the diagnostic performance of the Whooley depression questions and the EPDS during pregnancy and the early postnatal period, as well as the findings from the qualitative interviews with participants and health professionals and results of the cost-effectiveness analysis. Findings will be published in peer-reviewed journals and professional journals to ensure accessibility to health researchers and clinicians. Study findings will be published using the Standards for Reporting Diagnostic Accuracy studies (STARD) guidelines[54,

55]. We will present our findings at national conferences on perinatal depression, enabling the effective dissemination of our results to a wide target audience, to include midwives, health visitors, GPs and mental health professionals. We will also issue a press release to ensure coverage of our findings in the wider media. We will produce a short summary of the results for dissemination to all study participants as well as other relevant patient and other interest groups.

Contributors: EL, SA, PA, LD, CH, AK, RM, DMc and SG were responsible for the conception and design of the study and initial protocol. DM contributed to the development of the initial protocol. EL, LD, SGas, KS and BW contributed to refinements to the protocol and data acquisition. EL drafted the first draft of the manuscript. All authors contributed to and approved the final version of the manuscript.

Funding statement: This study was funded by the National Institute for Health Research Health Services and Delivery Research (NIHR HS&DR) programme (project number 11/2004/23). This paper presents independent research and the views and opinions expressed are those of the authors and do not necessarily reflect those of the HS&DR, NIHR, NHS or the Department of Health.

Competing Interests: None declared.

Patient consent: Obtained.

Ethics approval: North East – York Research Ethics Committee (Reference: 11/NE/0022).

REFERENCES

1. Murray C, Lopez A. *The global burden of disease: a comprehensive assessment of mortality and disability from disease, injuries andn risk factors in 1990*. Boston Mass: Harvard School of Public Health on behalf of the World Bank, 1996.
2. National Institute for Health and Care Excellence. *Antenatal and Postnatal mental health: Clinical management and service guidance. NICE clinical guideline 45.* London: The British Psychological Society & The Royal College of Psychiatrists, 2007.
3. National Institute for Health and Care Excellence. *Antenatal and postnatal mental health: Clinical management and service guidance. NICE clinical guideline 192.* London: The British Psychological Society & The Royal College of Psychiatrists, 2014.
4. Bennett HA, Einarson A, Taddio A, et al. Prevalence of depression during pregnancy: systematic review. *Obstet Gynecol* 2004;**103**(4):698-709
5. Marcus SM. Depression during pregnancy: rates, risks and consequences. *The Canadian journal of clinical pharmacology* 2009;**16**(1):e15-22
6. Gavin NI, Gaynes BN, Lohr KN, et al. Perinatal depression: a systmatic review of prevalence and incidence. *Obstet Gynecol* 2005;**106**:1071-83.
7. Boath E, Pryce A, Cox J. Postnatal depression: the impact on the family. *Journal of Reproductive and infant psychology* 1998;**16**(2-3):199-203
8. Murray L, Fiori-Cowley A, Hooper R, et al. The impact of postnatal depression and associated adversity on early mother-infant interactions and later infant outcome. *Child Dev* 1996;**67**(5):2512-26
9. Lovestone S, Kumar R. Postnatal psychiatric illness: the impact on partners. *Br J Psychiatry* 1993;**163**(2):210-16
10. Murray L, Sinclair D, Cooper P, et al. The socioemotional development of 5-year-old children of postnatally depressed mothers. *J Child Psychol Psychiatry* 1999;**40**(08):1259-71
11. Cogill S, Caplan H, Alexandra H, et al. Impact of maternal postnatal depression on cognitive development of young children. *Br Med J (Clin Res Ed)* 1986;**292**(6529):1165
12. Hearn G, Iliff A, Jones I, et al. Postnatal depression in the community. *Br J Gen Pract* 1998;**48**(428):1064
13. Secretary of State for Health. *National Service Framework - Mental Health*. London: HMSO, 1999.
14. Cox J, Holden J, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987;**150**(6):782-86
15. Shakespeare J. *Evaluation of screening for postnatal depression against the NSC handbook criteria*. London (UK): National Screening Committee, 2001
16. Hill C. *An evaluation of screening for postnatal depression against NSC criteria*. London (UK): National Screening Committee, 2010.
17. Henshaw C, Elliott S. *Screening for perinatal depression*. London (UK): Jessica Kingsley Publishers, 2005.
18. Hewitt CE, 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**(36)
19. 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**(7):439-45
20. 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**(02):465-86
21. Hewitt CE, Gilbody SM, Mann R, et al. Instruments to identify post-natal depression: Which methods have been the most extensively validated, in what setting and in which language? *Int J Psychiatry Clin Pract* 2010;**14**(1):72-76
22. Mann R, Gilbody S. Validity of two case finding questions to detect postnatal depression: A review of diagnostic test accuracy. *J Affect Disord* 2011;**133**(3):388-97

23. Mann R, Adamson J, Gilbody SM. Diagnostic accuracy of case-finding questions to identify perinatal depression. *CMAJ* 2012;**184**(8):E424-E30
24. Brealey SD, Hewitt C, Green JM, et al. Screening for postnatal depression–is it acceptable to women and healthcare professionals? A systematic review and meta-synthesis. *Journal of Reproductive and infant psychology* 2010;**28**(4):328-44
25. Mann R, Adamson J, Gilbody S. The acceptability of case-finding questions to identify perinatal depression. *British Journal of Midwifery* 2015;**23**(9)
26. Darwin Z, McGowan L, Edozien LC. Identification of women at risk of depression in pregnancy: using women's accounts to understand the poor specificity of the Whooley and Arroll case finding questions in clinical practice. *Arch Women Ment Health* 2015 doi: 10.1007/s00737-015-0508-1
27. Heron J, O'Connor TG, Evans J, et al. The course of anxiety and depression through pregnancy and the postpartum in a community sample. *J Affect Disord* 2004;**80**(1):65
28. National Institute for Health and Care Excellence. *Common mental health disorders: identification and pathways to care CG123*. London: National Institute for Health and Care Excellence, 2011.
29. Gilbody S, Sheldon T, Wessely S. Health policy: Should we screen for depression? *BMJ* 2006;**332**(7548):1027
30. Paulden M, Palmer S, Hewitt C, et al. Screening for postnatal depression in primary care: cost effectiveness analysis. *BMJ* 2009;**339**
31. Arroll B, Goodyear-Smith F, Kerse N, et al. Effect of the addition of a “help” question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *BMJ* 2005;**331**(7521):884
32. Mohr D, Hart S, Julian L, et al. Screening for depression among patients with multiple sclerosis: two questions may be enough. *Mult Scler* 2007;**13**(2):215-19
33. McManus D, Pipkin SS, Whooley MA. Screening for depression in patients with coronary heart disease (data from the Heart and Soul Study). *Am J Cardiol* 2005;**96**(8):1076
34. Mallen CD, Peat G. Screening older people with musculoskeletal pain for depressive symptoms in primary care. *Br J General Pract* 2008;**58**(555):688
35. Gaynes BN, Gavin N, Meltzer-Brody S, et al. Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes. *Evidence Report/Technology Assessment (Summary)*, 2005;1-8
36. Murray D, Cox JL. Screening for depression during pregnancy with the Edinburgh Depression Scale (EDDS). *Journal of Reproductive and infant psychology* 1990;**8**(2):99-107
37. Gruenberg AM, Goldstein RD, Pincus HA. *Classification of depression: research and diagnostic criteria: DSM-IV and ICD-10*. Weinheim: Wiley, 2005.
38. Singleton N, Bumpstead R, O'Brien M, et al. *Psychiatric morbidity among adults living in private households, 2000*. London, UK: The Stationery Office, 2001.
39. Evans M, Kessler D, Lewis G, et al. Assessing mental health in primary care research using standardized scales: can it be carried out over the telephone? *Psychol Med* 2004;**34**(1):157-62
40. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999;**282**(18):1737-44
41. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;**166**(10):1092
42. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med* 2002;**64**(2):258-66
43. Ware Jr JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;**34**(3):220-33
44. EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**(3):199-208
45. Gemmill AW, Leigh B, Ericksen J, et al. A survey of the clinical acceptability of screening for postnatal depression in depressed and non-depressed women. *BMC Public Health* 2006;**6**(1):211

46. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005;**58**(8):859
47. Tourangeau R. Cognitive science and survey methods. In: Jabine TB, Straf ML, Tanur JM, Tourangeau R, eds. *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press 1984:73-100.
48. Moustakas C. *Phenomenological research methods*: London: Sage Publications, Inc. 1994.
49. Silverman D. *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction* (Third edition). London: SAFE Publications Ltd, 2006
50. Balls P. Phenomenology in nursing research: methodology, interviewing and transcribing. *Nurs Times* 2009;**105**(32-33):30
51. Briggs A, Claxton C, Sculpher MJ. *Decision modelling for health economic evaluation*: Oxford University Press, 2006.
52. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ* 2005;**14**(4):339-47
53. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001;**10**(8):779-87
54. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;**351:h5527** doi: http://dx.doi.org/10.1136/bmj.h5527
55. Bossuyt PM, Reitsma JB, E Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;**326:41** doi: http://dx.doi.org/10.1136/bmj.326.7379.41

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4-8
	4	Study objectives and hypotheses	8
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	8
Participants	**6**	Eligibility criteria	9
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	9
	8	Where and when potentially eligible participants were identified (setting, location and dates)	9
	9	Whether participants formed a consecutive, random or convenience series	9
Test methods	**10a**	Index test, in sufficient detail to allow replication	10
	10b	Reference standard, in sufficient detail to allow replication	11
	11	Rationale for choosing the reference standard (if alternatives exist)	11
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	10-11
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	11-12
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	11-12
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	14
	15	How indeterminate index test or reference standard results were handled	14
	16	How missing data on the index test and reference standard were handled	14
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	
	18	Intended sample size and how it was determined	13
RESULTS			
Participants	**19**	Flow of participants, using a diagram	N/A
	20	Baseline demographic and clinical characteristics of participants	N/A
	21a	Distribution of severity of disease in those with the target condition	N/A
	21b	Distribution of alternative diagnoses in those without the target condition	N/A
	22	Time interval and any clinical interventions between index test and reference standard	N/A
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	N/A
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	N/A
	25	Any adverse events from performing the index test or the reference standard	N/A
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	N/A
	27	Implications for practice, including the intended use and clinical role of the index test	N/A
OTHER INFORMATION			
	28	Registration number and name of registry	N/A
	29	Where the full study protocol can be accessed	
	30	Sources of funding and other support; role of funders	19

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.