



King's Research Portal

DOI:

[10.1192/bjp.2017.9](https://doi.org/10.1192/bjp.2017.9)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Howard, L. M., Ryan, E. G., Trevillion, K., Anderson, F., Bick, D., Bye, A., Byford, S., O'Connor, S., Sands, P., Demilew, J., Milgrom, J., & Pickles, A. (2018). Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. *British Journal of Psychiatry*, 212(1), 50-56. <https://doi.org/10.1192/bjp.2017.9>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Title: The accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy.

^{1,2,3}Prof Louise Michele Howard (corresponding author)

¹Section of Women's Mental Health, PO31 Institute of Psychiatry, Psychology and Neuroscience;
King's College London

De Crespigny Park, London SE5 8AF

& ²South London and Maudsley NHS Foundation Trust; ³Womens' Health Academic Centre, King's College London

Telephone no: 02078488498; email louise.howard@kcl.ac.uk

⁴Dr Elizabeth G Ryan

¹Dr Kylee Trevillion

¹Fraser Anderson

^{3,6}Prof Debra Bick

¹Amanda Bye

⁵Prof Sarah Byford

¹Sheila O'Connor

¹Polly Sands

^{3,7}Jill Demilew

⁸Prof Jeannette Milgrom

⁴ Prof Andrew Pickles

⁴Biostatistics and Health Informatics Department, Institute of Psychiatry, Psychology and Neuroscience, King's College London

⁵King's Health Economics, Institute of Psychiatry, Psychology and Neuroscience, King's College London

⁶Florence Nightingale Faculty of Nursing and Midwifery, King's College London

⁷Women's Health, King's College Hospital NHS Foundation Trust, Denmark Hill, London SE5 8RS

⁶Parent-Infant Research Institute (PIRI®), Austin Health

Melbourne School of Psychological Sciences, University of Melbourne, Australia

Word count: 3785

Abstract

Background: There is limited evidence on the prevalence and identification of antenatal mental disorders.

Aims: To investigate 1) prevalence of mental disorders in early pregnancy; 2) diagnostic accuracy of depression-screening (Whooley) questions, compared with the Edinburgh Postnatal Depression Scale (EPDS), against the Structured Clinical Interview DSM-IV-TR.

Method: Cross-sectional survey of women responding to Whooley questions asked at first antenatal appointment. Women responding positively and a random sample of women responding negatively were invited to participate.

Results: Population prevalence was 27% (95%CI 22-32%): 11% (95%CI 8-14%) depression, 15% (95%CI 11-19%) anxiety disorders, 2% (95%CI 1-4%) obsessive-compulsive disorder, 0.8% (95%CI 0-1%) PTSD, 2% (95%CI 0.4-3%) eating disorders, 0.03% (95%CI 0-0.2%) bipolar disorder, 0.7% (95%CI 0-1%) borderline personality disorder. For identification of depression, likelihood ratios were 8.2 (Whooley) and 9.8 (EPDS). Diagnostic accuracy was similar in identifying any disorder (likelihood ratios 5.8 and 6).

Conclusions: Endorsement of Whooley questions in pregnancy indicates the need for a clinical assessment of diagnosis and could be implemented when maternity professionals have been appropriately trained on how to ask the questions sensitively, in settings where a clear referral and care pathway is available.

Declaration of interest: This paper summarises independent research funded by the National Institute for Health Research (NIHR) under the Programme Grants for Applied Research programme (ESMI Programme: grant reference number RP-PG-1210-12002) and the National Institute for Health Research (NIHR) / Wellcome Trust Kings Clinical Research Facility and the NIHR Biomedical Research Centre and Dementia Unit at South London and Maudsley NHS Foundation Trust and Kings College London. DB is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust (NIHR, CLAHRC-2013-10022). LMH is also supported by a National Institute for Health Research (NIHR) Research Professorship (NIHR-RP-R3-12-011). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The study team acknowledges the study delivery support given by the South London Clinical Research Network. Louise Howard, lead author, chaired the NICE CG192 guidelines development group on antenatal and postnatal mental health in 2012-4.

Introduction

Mental disorders during pregnancy are common (1) and are associated with adverse outcomes for women, pregnancy, the fetus, infant, childhood and adolescence (2-5). Recent clinical guidance(6) highlights the importance of identifying depression and other mental disorders early in pregnancy, and subsequently throughout the perinatal period, to facilitate early treatment and thus potentially mitigate subsequent adverse outcomes. The emerging evidence on the prevalence and impact of perinatal mental disorders across the diagnostic spectrum suggests that the ideal tool for case identification would indicate whether depression and other mental disorders may be present.

Most research has focused on identification of perinatal depression, with a recent evidence review(7) recommending the Edinburgh Postnatal Depression Scale (EPDS)(8), based on sensitivities of around 0.8 and specificities of 0.87, though values varied depending on the characteristics (e.g. ethnicity, socio-economic status) of the study population. No studies in this review, or subsequently to our knowledge, have systematically investigated the impact of these covariates on the diagnostic accuracy of the EPDS.

The review(7) also examined the accuracy of the PHQ-2,- a two-item scale, rated using a likert scale, or a yes/no response(9) (the latter sometimes known as the Whooley questions). Evidence on the usefulness of the Whooley questions as used in clinical practice was limited (8,9). Only one study(10) recruited women in early pregnancy and examined the PHQ-2, reporting an optimal cut point of 4 (specificity 0.79, sensitivity of 0.62) in a cohort of 213 women (13 of whom met criteria for major depressive disorder) recruited via advertisements in obstetric clinics(10). No data were available on use of these questions in the simpler yes/no binary format in early pregnancy. A systematic review of the Whooley questions across other settings (for men and women) reported a pooled sensitivity of 0.95 (95%CI 0.88 to 0.97) and pooled specificity of 0.65 (95%CI 0.55 to 0.74)(11). The only study of pregnant women in this review was a study of 126 women attending a UK maternity clinic at around 26–28 weeks' gestation who were given a self-administered questionnaire which included the Whooley questions and reported a prevalence of (minor and major) depression of 13.5% (95% CI 8.3%–21%); sensitivity was 100%(12).

In view of the limited evidence on the diagnostic accuracy of available tools as used in clinical practice, we aimed to investigate the diagnostic accuracy of the Whooley questions at the first antenatal appointment (“booking” appointment) in identifying 1) current depressive disorder, and 2) any disorder, compared with the EPDS, using a “gold standard” diagnostic instrument (the

Structured Clinical Interview DSM-IV(13)). We also aimed to examine the impact of relevant covariates on the performance of the EPDS instrument and to estimate the prevalence of mental disorders at antenatal booking.

Methods

Study design: Cross-sectional survey using a sampling design stratified according to being positive or negative (saying yes or no respectively) on either Whooley questions (“During the past month have you often been bothered by feeling down, depressed, or hopeless?”; “During the past month have you often been bothered by having little interest or pleasure in doing things?”), inviting a random sample of Whooley negative (W-) and all Whooley positive (W+) women to participate.

Study setting and population: We recruited women attending their booking appointment at an inner city maternity service in South-east London. Exclusion criteria were age <16, no response to the Whooley questions recorded, a previous comprehensive maternity booking elsewhere in the UK and a termination or miscarriage between the booking appointment and research interview.

Study procedures: A study advertisement was included in the pre-booking information pack sent to all women in advance of their appointment. On-line audit-trailed randomisation for enrolment into the study was carried out by trained researchers (research midwives and post-graduate research psychologists), once Whooley status (“positive”/“negative”) had been recorded by midwives. To enhance recruitment, researchers worked evenings and weekends to fit in with women’s work and childcare commitments. Recruitment and data collection were carried out once Whooley status (the index test) was determined, and the reference test (the “gold standard” diagnostic interview) was performed in those who consented to participate in the study. Usual care was delivered which could include a referral to primary or secondary mental health care services or a request for the primary care doctor to assess the woman’s mental health and whether she needs referral for treatment, depending on other aspects of the assessment carried out by the midwives.

Women who were W- were randomly selected - initially 1:4 and then 1:6 (see sample size calculation below). All potentially eligible W+ women and the randomly selected W- women were approached by a researcher (either on the day or, if not seen then, contacted by mobile telephone/ email/letter), who explained the study and invited women to consider participation. Interpreters were used to explain the study and translate where needed. Researchers interviewed women within a maximum

of 3 weeks of their antenatal booking appointment i.e. before any referrals for treatment made by midwives would have had a significant impact on mental health. Written informed consent was obtained.

Measures: The Whooley questions and the response to the additional “help” question (asked in women who respond yes to the questions, to identify those women who feel they want help), and socio-demographic data are recorded by the maternity staff. At interview the following instruments were administered by researchers:

1. Edinburgh Postnatal Depression Scale (EPDS)(8), a ten-item self-administered tool, administered by iPad where available and preferred (n=95), or pen and paper, using relevant language-specific tool
2. Structured Clinical Interview DSM-IV-TR Axis I Disorders (SCID-I-Research Version)(13) Axis I Mood Episodes, Mood Disorders and Anxiety Disorders module; SCID Axis I Eating Disorders module (SCID-I) and SCID II Personality Disorders sub-section module for borderline personality disorder(14). Researchers were trained to use the SCID, a “gold standard” diagnostic interview, over a three month period, and then met weekly with LMH to achieve consensus on diagnosis. Clinical information and Whooley status was available to these consensus meetings, though the two item responses as responded to when they are asked within the SCID (rather than the original response to the midwives) were used when assessing diagnostic criteria. Diagnosis of major depressive disorder included mild/moderate/severe depressive episode and mixed anxiety depression; women with bipolar disorder were classified as having current bipolar disorder (no women who were diagnosed with bipolar disorder were experiencing a depressive episode)

All women were offered information about sources of help and support (e.g. domestic violence, smoking and substance misuse services). Where the diagnostic interview identified a woman as having a mental disorder their midwife was informed, if the woman consented to this. Detailed standard operating procedures were used for this and other related studies(15) including contacting a senior clinician for concerns about safety, child protection and other potential sources of harm, and when to potentially breach confidentiality; in practice no breaches of confidentiality were necessary.

Sample size calculation: Power calculation for the two-phase design was undertaken using simulation with bootstrap estimation of CIs for the weighted estimators of sensitivity, specificity and

prevalence that corrected for the sample stratification. We assumed an overall prevalence of 9% depression and Whooley sensitivity of 0.95 and specificity of 0.89. Screening 6000 women by midwives, 66% of whom consent to participate, and sampling 54% of the W+ women (i.e. n=400) and 6% of the W- (i.e. n=200), would provide 600 women for interview; we expected 185 to be depressed. Assuming a sensitivity of 0.80 and specificity of 0.71, the width of the 95% CI for the EPDS sensitivity would be 0.19 and that for specificity 0.13. A conservative estimate of power based on the 185 disease cases only would have >90% power for a 0.8 v 0.65 sensitivity and specificity difference (comparing Whooley and the EPDS). As data collection was monitored, it became clear that adjustments to sampling fractions were necessary as there were fewer W+ women being recruited than anticipated while the original recruitment target of 200 W- women had been reached. After discussion with the independent Data Monitoring and Ethics Committee, we then aimed for 300 W+ and 300 W- women so that the two arms were recruited over the same time period, with random sampling of W- of 1:6.

Statistical Analysis

In all analyses (apart from examining differences in sociodemographic variables between W+ and W- women), sampling weights were used to account for the bias induced by the stratified sampling. Weights were based on the number of W+ and W- women in the study, out of all those that had maternity appointment bookings at the maternity unit during the study period (the sampling frame) (see Figure 1); the weights were 906/287 for W+ and 9057/258 for W-.

As pre-specified in our original grant application, for both the Whooley and the EPDS, we ascertained the rates of “true” and “false” positives and “true” and “false” negatives for: 1) depressive symptoms and 2) symptoms of any mental disorder. Sensitivity, specificity, positive predictive values (PPV), negative predictive values (NPV) and likelihood ratios were calculated. Appropriate cut-off scores for the EPDS were identified using receiver operating characteristics (ROC) analysis. This was implemented using the approach outlined by Pepe(16, 17) and the *rocrcg* command in Stata (v14.0(18)). The optimal cut-off point to discriminate between states (e.g. depressed and not depressed) was chosen based on sensitivity, specificity, PPV and NPV. We also estimated prevalence rates of disorders based on the weighted diagnostic interview responses. Bootstrap re-sampling of the weighted estimators was used for calculation of confidence intervals (other than for the prevalences of each of the categories in the severity of SCID depression (an ordinal outcome) for which we used the confidence interval estimates generated from Stata’s *svy* command).

As some covariates could affect the inherent discriminatory accuracy of the EPDS, we explored the effect of incorporating sociodemographic variables into the ROC curves for the EPDS using the approach described by *Janes et al*(19). It was assumed that the covariates affected the ROC curve only, and not the distribution among controls, and so the *rocreg* command was used in Stata (v14.0(18)) with the *roccov* option.

Missing data

Data for the help question were missing on 6 women who had responded yes to one of the two Whooley questions. Five women had completely missing data for EPDS items and were not included in EPDS analyses. Eleven women had 1-3 EPDS items missing. 521 (96%) women answered questions from all SCID modules but 21 (3.9%) declined the PTSD module (8 of whom had disclosed trauma during the interview) and other isolated non-completed modules occurred. A single round of predictive mean matching was performed using the *mi impute* function (predictive mean matching option) in Stata v14.0 to impute missing EPDS data for the 11 women who had 1-3 items (10-30%) missing. No imputation was performed for women who had more than 30% data missing in the EPDS items, which were treated as missing observations in relevant analyses (list-wise deletion performed in Stata). To account for missing observations in the SCID items, we used inverse probability weights that accounted for the Whooley sampling, as well as variables that were significant in predicting missingness of SCID responses (EPDS total score, ethnicity and employment status).

Ethical approval: The research was approved by the National Research Ethics Service, London Committee - Camberwell St Giles (ref no 14/LO/0075).

Results

Study sample

Between 10th November 2014 and 30th June 2016, 10004 women attended their initial antenatal appointment at the study site; 41 did not have a Whooley response recorded so the base population consisted of 9963 women. This base population was similar to the study population for age, ethnicity and number of children (see supplementary Table 1). Of the 882 Whooley negative (W-) women that were eligible, 624 (71%) did not participate and 478 (62%) of the 765 eligible Whooley positive (W+) did not participate (see Figure 1). Significant differences were found between the 287 W+ and the 258 W- participants, with W+ women more commonly being younger, single, living alone, having no formal educational qualifications/only high school qualifications, insecure immigration status and lower income (see Table 1).

Using weighted estimation, the population prevalence of a SCID disorder was 27% (95%CI 22 to 32%); with a prevalence of 11% (95%CI 8 to 14%) for depressive disorder (of which over half were mild depressive disorder (see Table 2), 15% (95%CI 11 to 19%) anxiety disorder, 2% (95%CI 1 to 4%) obsessive-compulsive disorder, 0.8% (95%CI 0 to 1%) PTSD, 2% (95%CI 0.4 to 3%) eating disorder, 0.03% (95%CI 0 to 0.2%) bipolar disorder and 0.7% (95%CI 0 to 1%) borderline personality disorder (see also Table 2).

No adverse events occurred from being asked the Whooley questions or taking part in the research interview. Health professionals were informed when severe disorders were identified and all participants consented to this information being shared with their midwife and/or GP.

Diagnostic accuracy of the Whooley questions for depression

SCID depression was found in 17 (6.6%) W- and 130 (45.3%) W+ women, where W+ was defined as answering yes to *either one* of the two questions. After adjustment for weighting, SCID depression was estimated to occur in 597 (7.6%) W- and 410 (45.3%) W+ women; no depression was found in 8460 (93.4%) W- and 496 (54.7%) W+ (see Table 3). Weighted sensitivity was 0.41, specificity 0.95, PPV 0.45, NPV 0.93, likelihood ratio (positive) 8.2, likelihood ratio (negative) 0.62 and area under the curve (AUC) for ROC curve 0.37 (95%CI 0.34 to 0.40). For W+ defined as answering yes to *either one* of the two Whooley questions *and* yes to the additional “help” question, sensitivity was 0.08 and specificity 0.99, with a PPV 0.66, NPV 0.83, likelihood ratio (positive) 8, likelihood ratio (negative) 0.93, and AUC for ROC curve 0.21 (95%CI 0.19 to 0.23). When W+ was defined as answering yes to *both* of the two Whooley questions, SCID depression was found in 67 (16.2%) W- and 80 (60.6%) W+

women. Weighted sensitivity was 0.25, specificity 0.98, with a PPV 0.61, NPV 0.84, likelihood ratio (positive) 12.5, likelihood ratio (negative) 0.77, and AUC of the ROC curve 0.24 (95%CI 0.21 to 0.26).

Diagnostic accuracy of the EPDS for depression

The range of EPDS scores was 0-28, with a median of 7 (IQR: 4-13). Using a cut-off of 12/13 (which was optimal for diagnostic accuracy - see Table 3), SCID depression was found in 49 (12.2%) EPDS- women and 98 (68.5%) EPDS+ women. This cut-off resulted in weighted sensitivity 0.59, specificity 0.94, PPV 0.52, NPV 0.95, likelihood ratio (positive) 9.8, likelihood ratio (negative) 0.44, and AUC for the ROC curve of 0.89 (95%CI 0.88 to 0.90) (see supplementary table 2 for 2x2 tables of weighted prevalences). ROC curve analysis found no evidence of a difference in the ability of the EPDS to discriminate between cases and non-cases among the five ethnicity categories ($\chi^2(4)=3.52$; $p=0.48$), the five income categories ($\chi^2(4)=6.89$; $p=0.14$), the three education categories ($\chi^2(2)=2.48$; $p=0.29$) nor for those who used an iPad rather than paper for completion ($\chi^2(1)=1.48$; $p=0.22$). However, there was evidence that the discriminatory ability of the EPDS decreased with increasing age analysed as a continuous variable ($\chi^2(1)=19.12$; $p<0.0001$) (see Figure 2). The EPDS also performed better in those who participated in the study using an interpreter ($n=40$) ($\chi^2(1)=5.45$; $p=0.02$).

Diagnostic accuracy for any disorder

A SCID disorder was found in 242 (45%) participants (27% had depressive disorders and 21% had anxiety disorders) and 294 (54%) had no SCID diagnosis. Using weighted estimation, the population prevalence of any disorder was 27% (95%CI 22 to 32%).

Whooley questions: A SCID disorder was found in 55 (21.9%) W- women and 187 (65.6%) W+ women, where W+ was defined as answering yes to *either one* of the two questions. Weighted sensitivity (see supplementary table 3 for 2x2 table of weighted prevalences) was 0.23, specificity 0.96, PPV 0.66, NPV 0.78, likelihood ratio (positive) 5.8, likelihood ratio (negative) 0.80 and AUC for ROC curve 0.21 (95%CI 0.2 to 0.23). When the “help” question was added, sensitivity was 0.05, specificity 1.00, PPV 0.86, NPV 0.65, negative likelihood ratio 0.95 and AUC for ROC curve was 0.11 (95% CI 0.10 to 0.12). When W+ was defined as answering yes to *both* questions the sensitivity was 0.06, specificity 0.99, PPV 0.78, NPV 0.66, likelihood ratio (positive) 6, likelihood ratio (negative) 0.95, and AUC for the ROC curve was 0.12 (95% CI 0.11 to 0.13).

EPDS: Using a cut off of 12/13, 121 (30.7%) *EPDS*- women and 121 (85.2%) *EPDS*+ women had a SCID disorder (see Supplementary Table 3 for 2x2 table of weighted prevalences). Weighted sensitivity was 0.3, specificity 0.95, PPV 0.67, NPV 0.79, likelihood ratio (positive) 6, likelihood ratio 0.74 (negative), and AUC for ROC curve 0.74 (95%CI 0.73 to 0.75).

Discussion

Main Findings

The ten-item Edinburgh Postnatal Depression Scale (EPDS) performed better in correctly identifying major depression (likelihood ratio 9.8) than the two item Whooley questions (likelihood ratio 8.2) in early pregnancy. However the difference in diagnostic accuracy was not large, and both tools had high specificity. There have been no comparable studies in early pregnancy but our study, which aimed to validate the use of the Whooley questions when routinely asked face-to-face by midwives, suggests that the Whooley has a lower sensitivity for identification of depression than often reported in other studies. This may be due to different methods of administration of the questions, as previous studies have not usually validated the Whooley as used in maternity practice, but rather have used a written format administered by a researcher (11). However, the low sensitivity of the Whooley may be also partly due to the variation in how the questions were asked in clinical practice by staff who had not usually been trained in perinatal mental health. The sensitivity of the EPDS was comparable to some studies though others have reported higher sensitivity in pregnancy(7). The (generally) lower sensitivity of the screening tools found in our study compared with others, may also reflect the larger, more representative study population included here (which in this study included women from very diverse backgrounds and those who did not speak English) and, for the Whooley, the delay between the midwives asking the questions and the diagnostic interview being administered.

The administration of the 10 questions of the EPDS is, however, potentially burdensome in busy maternity settings and may outweigh the slightly improved diagnostic accuracy of the EPDS when considering how to identify antenatal depression in routine maternity care. It is therefore useful to note that use of iPads in administering the EPDS did not reduce its effectiveness. The EPDS, a self-complete tool, therefore could be completed by women when they are not being seen by a midwife e.g. while in the waiting room. It should be noted though, that we found that responses to the EPDS in older women were less discriminatory in identifying depression. We speculate that this may reflect a longer duration of mental disorders and associated anticipated discrimination (20), or self-perceived resilience.

In clinical practice, maternity professionals need to identify whether or not a woman has any mental disorder, not only depression. It is therefore particularly important that, in this study, there was little difference in diagnostic accuracy between the Whooley questions and the EPDS in identifying a mental disorder: both tools had low sensitivity (0.23 for the Whooley, 0.3 for the EPDS) and high specificity (0.94 and 0.95 respectively), with similar likelihood ratios. In practice, this means that

pregnant women presenting for their first antenatal appointment who have a mental disorder are 5.8 times more likely to say yes to one of the Whooley questions, (or 6 times more likely to score above 12 on the EPDS) than those without a mental disorder, supporting use of either instrument in routine practice; a positive screen then needs to be followed by a clinical assessment by an appropriate health practitioner to establish the clinical diagnosis and appropriate intervention(21).

Implications

Our findings confirm that the Whooley questions are a useful tool for case identification in early pregnancy in settings where face to face questions can be asked as part of a general discussion about health; a positive Whooley response suggests the respondent may have a mental disorder (not necessarily depression), and needs further clinical assessment. Questions about mental health that can be asked quickly and easily by midwives at routine planned contacts also indicate to pregnant women that this is a service that addresses mental and physical health; such questions, in the context of a supportive open discussion, also provide an opportunity to discuss a woman's replies in the context of her psychosocial circumstances. There is evidence that *how* the Whooley questions are asked by midwives determines their acceptability(22) and the Whooley questions should therefore only be implemented when midwives and obstetricians have been appropriately trained on how to ask the questions sensitively, in settings where a clear referral and care pathway is available.

This study does not provide direct evidence on whether midwives should routinely ask screening questions, as women were not randomised to routine enquiry to examine whether being asked improves health outcomes. However, there is evidence that unless mental health questions are asked routinely, women from some backgrounds (e.g. non-white groups) are less likely to be asked about mental health(23, 24). Routine enquiry could therefore address "the inverse care law" in relation to maternity care(23), and practitioners' unconscious biases. Standardised questions are sometimes viewed as a "tick box" exercise by staff and women(25), but it is striking that where women were asked the EPDS questions (i.e. when they needed an interpreter) there was significantly better discriminatory performance of the instrument suggesting that being asked questions face-to-face may facilitate disclosure of problems. A similar study, validating the questions when asked by relevant health professionals such as health visitors in the postpartum period, would be useful.

Strengths and Limitations

This study assessed the accuracy of the Whooley questions being asked by midwives at a routine maternity contact rather than validating responses to researchers. Other strengths include the use of a diagnostic interview; an efficient, well powered study design; and a diverse study population. Limitations include the relatively low response rate, the delay in administering the EPDS and the SCID after the initial booking appointment when the Whooley questions were asked, some missing data, and the use of a single maternity site in inner city London.

Conclusions

The two item Whooley questions can be asked routinely by midwives when women attend for their routine antenatal booking appointment and are a quick method for identifying that a mental disorder may be present. This study also supports an alternative strategy of a self-complete EPDS, using a tablet or paper. A positive screen will necessitate further comprehensive psychosocial assessment for identification of the type and severity of mental disorder and related problems, and subsequent treatment to reduce maternal and fetal morbidity.

Acknowledgements: We gratefully acknowledge the advice received from our Patient and Public Advisory Group (Clare Dolman, Sarah Spring, Ceri Rose, Liberty Mosse, Amanda Grey, Henry Fay, Kathryn Grant, Maria Bavetta, Eleanor O’Sullivan, Jesse Hunt, Diana Rose, chair), our Programme Steering Committee (Prof Rona McCandlish (Chair), Dr Heather O’Mahen, Dr Pauline Slade, Ceri Rose, Sarah Spring, and Rosemary Jones) and our Data Monitoring and Ethics Committee (Roch Cantwell (chair), Liz McDonald-Clifford, Marian Knight, Stephen Bremner). We also want to take the opportunity to thank the women who participated in this study.

Figure 1: Flow chart of women through the study

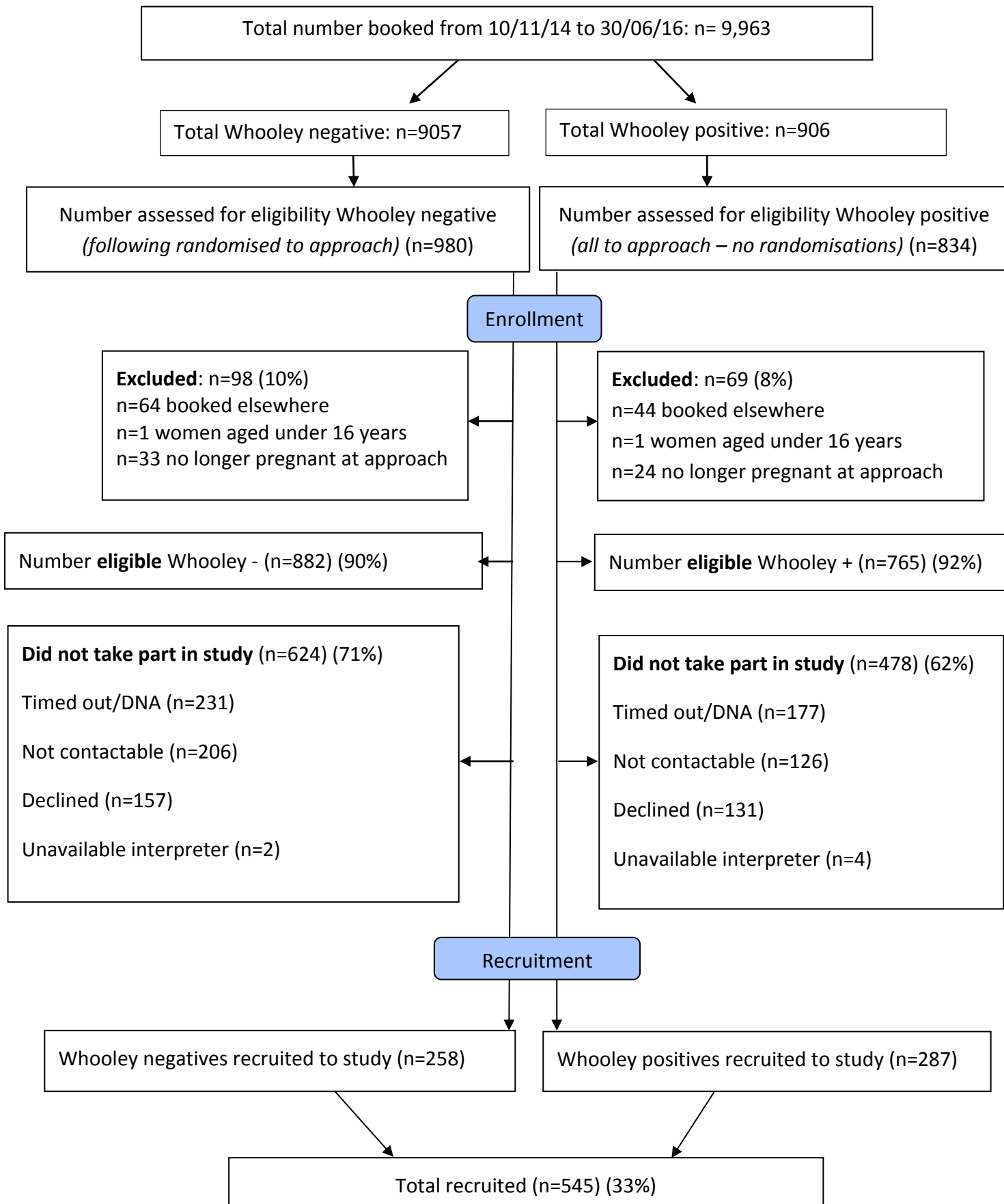


Figure 2: Receiver operating characteristic (ROC) curves of EPDS with covariate adjustment

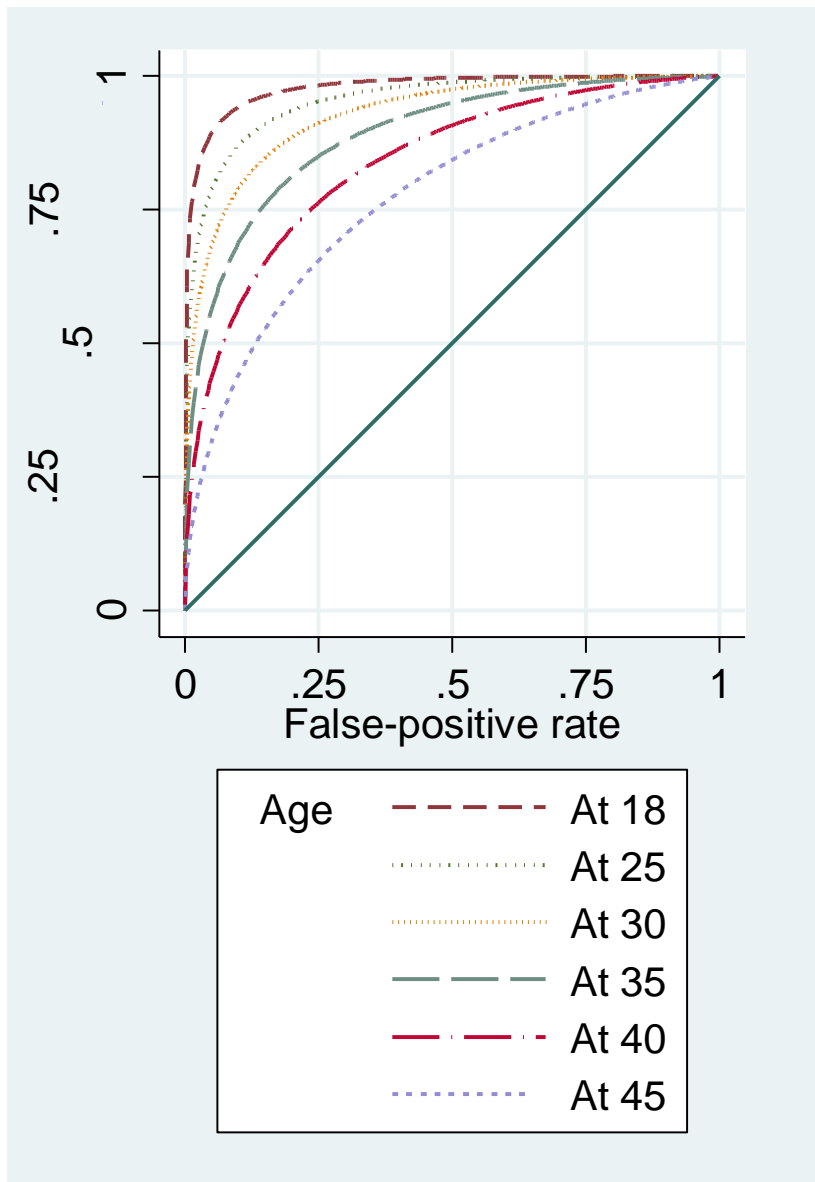


Table 1: Characteristics of the study population by Whooley and EPDS status

	Whooley Positive	Whooley Negative	P value	EPDS* Positive	EPDS Negative	P value	Total
Age							
<25	45 (16%)	12 (5%)	P<0.001	26 (18%)	31 (8%)	P=0.002	57 (10%)
25-29	53 (18%)	48 (19%)		29 (20%)	72 (18%)		101 (19%)
30-39	165 (57%)	176 (68%)		80 (56%)	261 (65%)		341 (63%)
40+	24 (8%)	22 (9%)		8 (6%)	38 (9%)		46 (8%)
Ethnicity							
White	140 (49%)	144 (56%)	P=0.460	63 (44%)	221 (55%)	P=0.156	284 (52%)
Black	99 (34%)	78 (30%)		58 (41%)	119 (30%)		177 (32%)
Asian	14 (5%)	11 (4%)		7 (5%)	18 (5%)		25 (5%)
Mixed	15 (5%)	8 (3%)		5 (4%)	18 (5%)		23 (4%)
Other	19 (7%)	17 (7%)		10 (7%)	26 (6%)		36 (7%)
Partner status							
Partner not cohabiting	59 (21%)	23 (9%)	P<0.001	29 (20%)	53 (13%)	P<0.001	82 (15%)
Cohabiting/ married	178 (62%)	214 (83%)		80 (56%)	312 (78%)		392(72%)
Single	50 (17%)	21 (8%)		34 (24%)	37 (9%)		71 (13%)
Living status							
Alone	46 (16%)	25 (10%)	P<0.001	30 (21%)	41 (10%)	P<0.001	71 (13%)
Spouse/partner	174 (61%)	202 (79%)		79 (56%)	297 (74%)		376 (70%)
Parents/family	29 (10%)	17 (7%)		13 (9%)	33 (8%)		46 (9%)
Friends	10 (4%)	4 (2%)		8 (6%)	6 (2%)		14 (3%)
Other	25 (9%)	9 (4%)		12 (8%)	22 (6%)		34 (6%)
Employment status							
Working	175 (61%)	180 (70%)	P=0.086	83 (58%)	272 (68%)	P=0.051	355 (65%)
Student	8 (3%)	8 (3%)		5 (4%)	11 (3%)		16 (3%)
Unemployed	41 (14%)	23 (9%)		19 (13%)	45 (11%)		64 (12%)
Homemaker	38 (13%)	36 (14%)		19 (13%)	55 (14%)		74 (14%)
Not working due to illness/Other	23 (8%)	11 (4%)		16 (11%)	18 (4%)		34 (6%)
Immigration status							
UK National	168 (63%)	162 (63%)	P=0.010	77 (54%)	253 (63%)	P = 0.023	330 (61%)
EEA citizen	39 (14%)	38 (15%)		17 (12%)	60 (15%)		77 (14%)
Indefinite leave to remain/ Exceptional leave to remain	31 (11%)	24 (9%)		19 (13%)	36 (9%)		55 (10%)
Temporary admission/ Awaiting initial decision	34 (12%)	11 (4%)		20 (14%)	25 (6%)		45 (8%)
Other	15 (5%)	23 (9%)		10 (7%)	28 (7%)		38 (7%)
Needs an interpreter							
Yes	27 (9%)	13 (5%)	P = 0.051	13 (9%)	27 (7%)	P = 0.350	40 (7%)
Educational history							
No/only school qualifications	78 (27%)	42 (16%)	P=0.004	36 (25%)	84 (21%)	P=0.063	120 (22%)

Training/Higher Certificate/Diploma	77 (27%)	66 (26%)		45 (31%)	98 (24%)		143 (26%)
Degree/Postgraduate	132 (46%)	150 (58%)		62 (43%)	220 (55%)		282 (52%)
Other children							
None	147 (51%)	124 (48%)	P=0.716	69 (48%)	202 (50%)	P=0.813	271 (50%)
1	88 (31%)	87 (34%)		49 (34%)	126 (31%)		175 (32%)
≥2	52 (18%)	47 (18%)		25 (17%)	74 (18%)		99 (18%)
Booked late for antenatal care							
No	229 (80%)	221 (86%)	P=0.071	114 (80%)	336 (84%)	P=0.296	450 (83%)
Yes	58 (20%)	37 (14%)		29 (20%)	66 (16%)		95 (17%)
Income							
< £15000	51 (24%)	26 (13%)	P=0.002	39 (13%)	38 (34%)	P<0.001	77 (19%)
£15,000-£30,999	43 (20%)	28 (14%)		47 (15%)	24 (22%)		71 (17%)
£31,000-£45,999	27 (13%)	33 (16%)		47 (15%)	13 (12%)		60 (14%)
£46,000-£60,999	29 (14%)	34 (17%)		52 (17%)	11 (10%)		63 (15%)
£61,000 or more	60 (29%)	85 (41%)		120 (39%)	25 (23%)		145 (35%)

*cut-off of 13 is used as this was found to have a good compromise between the sensitivity and specificity.

Table 2: Population prevalence of diagnoses by Whooley and EPDS status (95%CI)

	Whooley Positive	Whooley Negative	EPDS* Positive	EPDS Negative	Prevalence
Major depression [^]	47% (40-53%)	7% (4-10%)	53% (37-71%)	5% (3-7%)	11% (8-14%)
<i>(No depression</i>	<i>56% (49-61%)</i>	<i>93% (89-96%)</i>	<i>48% (34-62%)</i>	<i>95% (93-97%)</i>	<i>90% (86-92%)</i>
<i>Mild depression</i>	<i>23% (18-29%)</i>	<i>4% (2-7%)</i>	<i>27% (16-41%)</i>	<i>3% (2-6%)</i>	<i>6% (4-9%)</i>
<i>Moderate depression</i>	<i>20% (16-25%)</i>	<i>3% (1-6%)</i>	<i>25% (15-38%)</i>	<i>2% (0.6-4%)</i>	<i>4% (3-8%)</i>
<i>Severe depression</i>	<i>1% (0.4-4%)</i>	<i>0%</i>	<i>1% (0.3-3%)</i>	<i>0%</i>	<i>0.1% (0-0.3%)</i>
<i>Mixed anxiety/depression)</i>	4% (2-6%)	0%	1% (0 -2%)	0.3% (0.1-0.4%)	0.4% (0.2-0.6%)
Any anxiety disorder	30% (24-36%)	13% (9-18%)	23% (12-35%)	14% (10-18%)	15% (11-19%)
Obsessive compulsive disorder	5% (3-7%)	2% (0.4-4%)	3% (1-5%)	2% (1-4%)	2% (1-4%)
Eating disorder	5% (2-8%)	1% (0-4%)	3% (1-5%)	1% (0.1-3%)	2% (0.4-3%)
PTSD	5% (2-7%)	0.4% (0.1-3%)	6% (0-11%)	0.1% (0-0.3%)	0.8% (0-1%)
Bipolar disorder I	0.3% (0.1-1%)	0%	0%	0.04% (0-0.1%)	0.03% (0-0.2%)
Bipolar disorder II	0.3% (0.1-1%)	0%	0%	0.04% (0-0.1%)	0.03% (0-0.2%)
Borderline personality disorder	4% (2-6%)	0.4% (0-3%)	2% (1-4%)	0.5% (0-1.3%)	0.7% (0-1%)
Any SCID	67% (60-73%)	22% (17-29%)	67% (51-84%)	21% (16-27%)	27% (22-32%)

[^] This includes major depressive disorder and mixed anxiety and depression. Minor depression rates were 10% (95% CI 7 to 14%) in Whooley positive; 2% (95%CI 0.8 to 5%) in Whooley negative; 9% (95%CI 4 to 22%) in EPDS positive; 2% (95%CI 0.8 to 4%) in Whooley negative; an overall prevalence of 3% (95%CI 1 to 5%). * cut off of 12/13 used for EPDS negative/positive.

Table 3: Performance of Edinburgh Postnatal Depression Scale for different cut-off values for depression

Cut-off	Sensitivity	Specificity	PPV*	NPV*
11	0.73	0.88	0.40	0.97
12	0.68	0.92	0.48	0.96
13	0.59	0.94	0.52	0.95
14	0.46	0.95	0.50	0.94
15	0.44	0.96	0.56	0.94

*PPV= Positive predictive value; NPV=Negative predictive value

References

1. Howard LM, Molyneaux E, Dennis C-L, Rochat T, Stein A, Milgrom J. Non-psychotic mental disorders in the perinatal period. *The Lancet*. 2014; 384(9956): 1775-88.
2. Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, et al. Effects of perinatal mental disorders on the fetus and child. *The Lancet*. 2014; 384(9956): 1800-19.
3. Garthus-Niegel S AS, Martini J, von Soest T, Eberhard-Gran M. The impact of postpartum post-traumatic stress disorder symptoms on child development: a population-based, 2-year follow-up study. *Psychol Med*. 2017; 47(1): 161-70.
4. Solmi F SH, Stahl D, Treasure J, Micali N. Low birth weight in the offspring of women with anorexia nervosa. *Epidemiol Rev*. 2014; 36: 49-56.
5. Micali N DSB, Ploubidis GB, Simonoff E, Treasure J. The effects of maternal eating disorders on offspring childhood and early adolescent psychiatric disorders. *Int J Eat Disord*. 2014; 47(4): 385-93.
6. National Institute for Health and Care Excellence. Antenatal and Postnatal Mental Health guidelines - Clinical guidelines CG192. National Institute for Health and Care Excellence (NICE), 2014.
7. O'Connor E, Rossom R, M H. Primary Care Screening for and Treatment of Depression in Pregnant and Postpartum Women: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 2016; 315(4): 388-406.
8. Cox JL, Holden J, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*. 1987; 150(6): 782-6.
9. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med*. 1997; 12(7): 439-45.
10. Smith M, Gotman N, Haikin L, Yonkers K. Do the PHQ-8 and the PHQ-2 accurately screen for depressive disorders in a sample of pregnant women? *Gen Hosp Psychiatry*. 2010; 32(5): 544-8.
11. Bosanquet K, Bailey K, Gilbody S, Harden M, Manea L, Nutbrown S, McMillan D et al. Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. *BMJ Open*. 2015; 5(12).
12. Mann R, Adamson J, Gilbody S. The acceptability of case-finding questions to identify perinatal depression. *British Journal of Midwifery*. 2015; 23(9): 630-8.
13. First MB, Spitzer RL, Gibbon M, Williams JB. Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition. New York: Biometrics Research, New York State Psychiatric Institute. 2002.
14. Zanarini MC, Frankenburg FR, Sickel AE, Yong L. The diagnostic interview for DSM-IV personality disorders (DIPD-IV). Belmont, MA: McLean Hospital. 1996; 340.
15. Trevillion K, Domoney J, Pickles A, Bick D, Byford S, Heslin M, et al. Depression: an exploratory parallel-group randomised controlled trial of Antenatal guided self help for Women (DAWN): study protocol for a randomised controlled trial. *Trials*. 2016; 17(503).
16. Pepe M. Receiver operating characteristic methodology. *Journal of the American Statistical Association*. 2000; 95: 308-11.
17. Pepe M. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, 2003.
18. StataCorp. *Stata Statistical Software: Release 14*. StataCorp LP, 2015.
19. Janes H, Longton G, Pepe M. Accommodating Covariates in ROC Analysis. *Stata Journal*. 2009; 9(1): 17-39.
20. Farrelly S CS, Gabbidon J, Jeffery D, Dockery L, Lassman F, Brohan E, Henderson RC, Williams P, Howard LM, Thornicroft G; MIRIAD study group. Anticipated and experienced discrimination amongst people with schizophrenia, bipolar disorder and major depressive disorder: a cross sectional study. *BMC Psychiatry*. 2014; 14(157).
21. Whooley M. Screening for Depression—A Tale of Two Questions. *JAMA Internal Medicine*. 2016; 176(4): 436-8.

22. 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. *Archives of women's mental health*. 2016; (19): 41-9.
23. Redshaw M HJ. Who is actually asked about their mental health in pregnancy and the postnatal period? Findings from a national survey. *BMC psychiatry*. 2016; 16(1): 322.
24. Prady SL, Pickett KE, Petherick ES, Gilbody S, Croudace T, Mason D, et al. Evaluation of ethnic disparities in detection of depression and anxiety in primary care during the maternal period: combined analysis of routine and cohort data. *The British Journal of Psychiatry*. 2016; 208(5): 453-61.
25. Darwin Z, McGowan L, Edozien LC. Antenatal mental health referrals: Review of local clinical practice and pregnant women's experiences in England. *Midwifery*. 2015; 31(3): e17-e22.