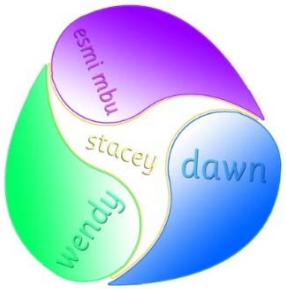


A cross-sectional study comparing the Whooley questions and Edinburgh postnatal depression scale against a diagnostic assessment in identifying depression in pregnancy

Dr Elizabeth Ryan on behalf of the ESMI team

Statistician, Biostatistics Department, IoPPN, King's College London

29 September 2016

NHS
National Institute for Health Research

Outline

- Background
  - Introduction
  - Aims
- Methods
  - Study design
  - Measures of Interest
  - Statistical Methods
- Results
  - Patient demographics
  - Sensitivity, specificity, PPV and NPV of Whooley
  - ROC curve analysis for EPDS
  - Comparison of EPDS and Whooley
- Discussion/Conclusions
- References

Background

Introduction

- Growing evidence that many women experience mental health problems during pregnancy
  →Depression most common disorder → approx. 9% prevalence during pregnancy (Gavin et al.,2005)
- Identification of depression cases very important as it may impact on relationships between the mother and partner/family, mother-baby interaction, infant growth, and on longer-term emotional and cognitive development of the baby

- Most studies on pregnancy prevalence of depression have focused on 3rd trimester
- Questions about depression are primarily asked at antenatal booking in UK (i.e., 1st trimester)
- Antenatal mental disorders are often unrecognised and untreated, despite frequent contact with healthcare professionals during pregnancy

The UK National Institute for Clinical Excellence (NICE) recommends maternity professionals use the **two Whooley questions** (Aroll et al., 2003) to identify depressive disorders in the perinatal and postnatal period.

During the past month have you often been bothered by :

- **Feeling down, depressed, or hopeless?**
- **Having little interest or pleasure in doing things?**

- **Yes to either question → Whooley positive**

- **Unclear whether these questions are the optimal method to do this**

- The **Edinburgh Postnatal Depression Scale** (EPDS (Cox et al., 1987)) is an alternative measure consisting of **10 self-reported items** that have been used extensively in primary care for the detection of depression in the perinatal period

EPDS Questionnaire (Cox et al., 1987)

1. I have been able to laugh and see the funny side of things
2. I have looked forward with enjoyment to things
3. I have blamed myself unnecessarily when things went wrong
4. I have been anxious or worried for no good reason
5. I have felt scared or panicky for no very good reason
6. Things have been getting on top of me
7. I have been so unhappy that I have had difficulty sleeping
8. I have felt sad or miserable
9. I have been so unhappy that I have been crying
10. The thought of harming myself has occurred to me

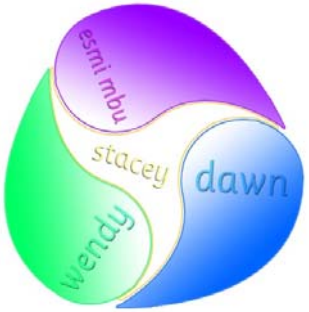
- Responses are on a 4 point scale (0-3), with different categories for each question.
- 2 of the items are reverse coded
- Scores range from 0 - 30

Aims

To investigate:

- The **sensitivity, specificity and positive predictive value** of the two Whooley questions (currently routinely used by midwives in England) compared with these values for the EPDS, using a gold standard diagnostic interview as a reference, for the identification of depression at antenatal booking (and subsequently later in pregnancy and post-delivery)
- Determine the appropriate **cut-off point** for the EPDS to classify someone as being "depressed" → 12, 13, or 14??
- Determine whether the **discriminatory accuracy** of the EPDS varies between different groups (e.g., ethnicity, country of origin, income, age etc)

Methods



Study Methods

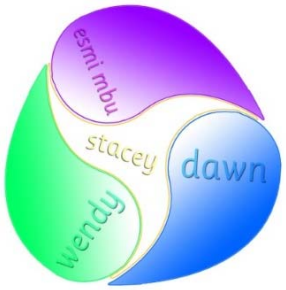
Study design:

- Cross-sectional study with a two-phase sampling design drawing a random sample stratified according to being positive or negative on the Whooley questions
- Nested cohort study for those who agree to continuing participation in the follow up study.

Study setting:

- Inner city maternity service (6000 births/year) at King's College Hospital, London
- Ethnically and socially diverse population.
- Study period: 10th November, 2014 to 30th June, 2016

NHS
National Institute for Health Research



Study population

Inclusion criteria:

- Women aged >15 who answered the Whooley questions at antenatal booking;

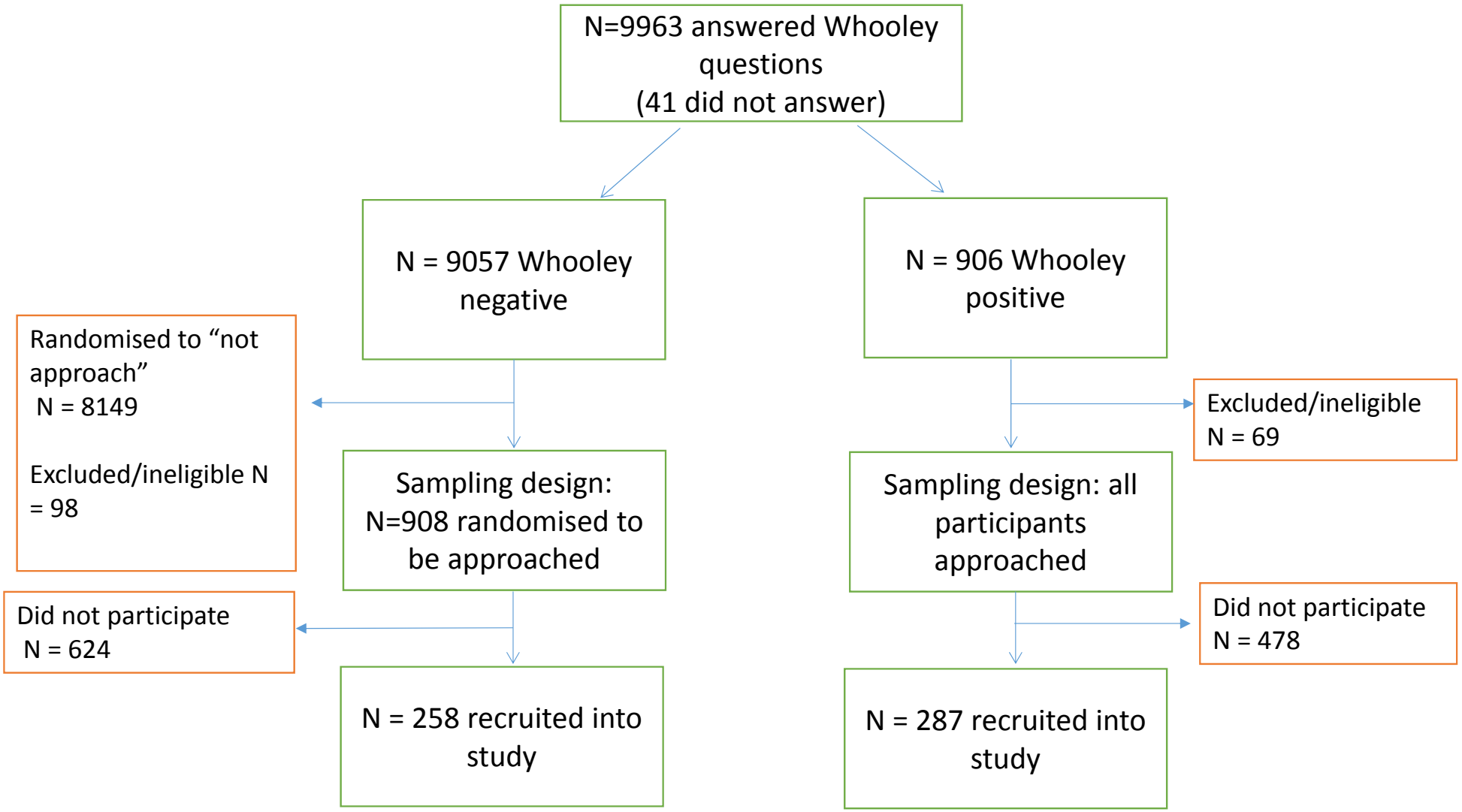
Exclusion criteria:

- Women >15 who lack mental capacity to provide informed consent;
- Women who decline answering the Whooley questions;
- Women who have already undergone a comprehensive maternity booking in the UK;
- Women who have had a termination or miscarriage between booking and baseline interview

Recruitment and sampling

- At booking, clinic midwives ask women the recommended Whooley questions – recorded via an electronic booking system.
- Whooley status ("positive"/"negative") determined by the clinic midwives.
- **¼ Whooley negative women were randomly sampled to be approached for inclusion into the study; all Whooley positive women were approached**
- Women who are selected for potential enrolment are approached by a research midwife
- Researchers will interview participants within 3 weeks of antenatal booking and administer the EPDS and Structured Clinical Interview DSM-IV (SCID) → Gold standard for diagnosing depression and other mental disorders

Study design



Measures of interest

- Whooley questionnaire
- Edinburgh Postnatal Depression Scale (EPDS)
- The Structured Clinical Interview DSM-IV Axis I Mood Episodes, Mood Disorders and Anxiety Disorders module (SCID questionnaire)

Whooley questionnaire

- Comprise of **two** items that measure symptoms of depression
- Also an additional help question “is this something you feel you need or want help with?”
- The Whooley and “help” questions are routinely asked at antenatal booking across English maternity services.
- Has previously been investigated using a total score from a Likert scale, but NICE recommends Whooley questions looked at as a **binary** measure

- Have demonstrated good sensitivity and specificity: 0.97 and 0.67, respectively with the CIDI (Arroll et al., 2003)
- This measure will be used at baseline (and follow-up assessments at 28 weeks pregnancy and three months post-delivery.)

EPDS

- **Ten item** self-administered screen for perinatal depression
- Total score ranges from 0-30
- Has a positive predictive value for antenatal major depression 60-80% (cut-off score 14/15) (Gibson et al., 2009)
- This measure will be used at baseline (and follow-up assessments at 28 weeks pregnancy and three months post-delivery.)

The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID)

- SCID (First et al., 2012; American Psychiatric Association, 2000) is a semi-structured diagnostic interview which has been widely used in psychiatric research
- The interview consists of standardised diagnostic questions arranged in modules corresponding to each DSM-IV Axis I disorder
- This research study will use only the Mood Episodes, Mood Disorders and Anxiety Disorders modules
- For our analyses, Depression = SCID major depressive disorder/major depressive episode/mixed anxiety depressed
- SCID is used as the "gold standard" for diagnosis of depression in our analyses

Statistical Methods

- Expansion weights used to account for oversampling of Whooley positives and undersampling of Whooley negatives
- **Weights calculated as: 906/287 (tve) and 9057/258 (nve)**

→ Didn't have any information available on those that didn't take part in the study (e.g., age, ethnicity, etc) to be able to use inverse probability weights from logistic regressions → Will have this information for future analyses

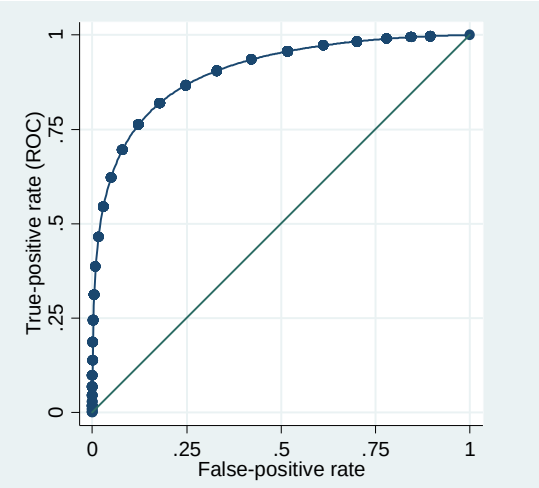
- Complete case analysis performed (EPDS available for 533/545 ≈98%)
- Can declare as survey data in Stata (svyset) and use weights in analyses (using svy command, or in options of Stata commands)
- ROC curves used to determine impact of altering threshold of EPDS on false tve/nve rate (use SCID diagnosis to determine "true" diagnosis)

ROC Curves

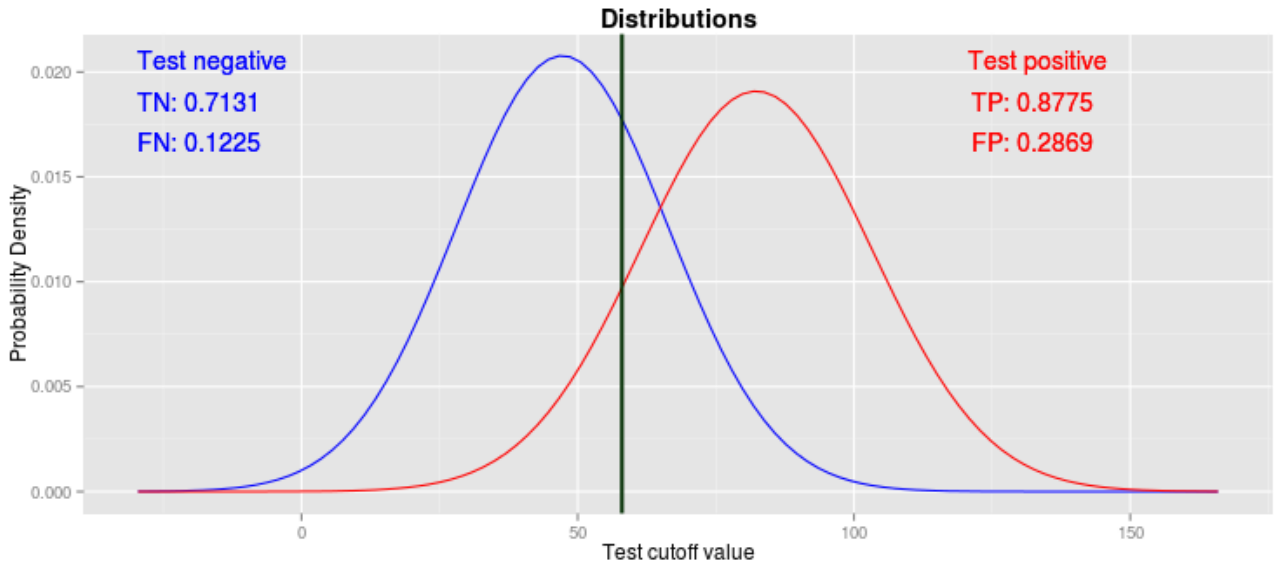
- Receiver operating characteristic (ROC) curves are used to determine the best cut-off value to distinguish between "cases" and "non-cases" from a clinical measure
  → ordinal (e.g., questionnaire) or continuous measures (e.g., enzyme concentrations)
  - Can also be used to determine the accuracy of a measure
  - The cut-off is varied throughout the whole range of possibilities and the sensitivity and specificity are estimated for each cut-off value
  - **The ROC curve is a plot of the sensitivity vs 1 – specificity**
    → True tve vs false tve rate

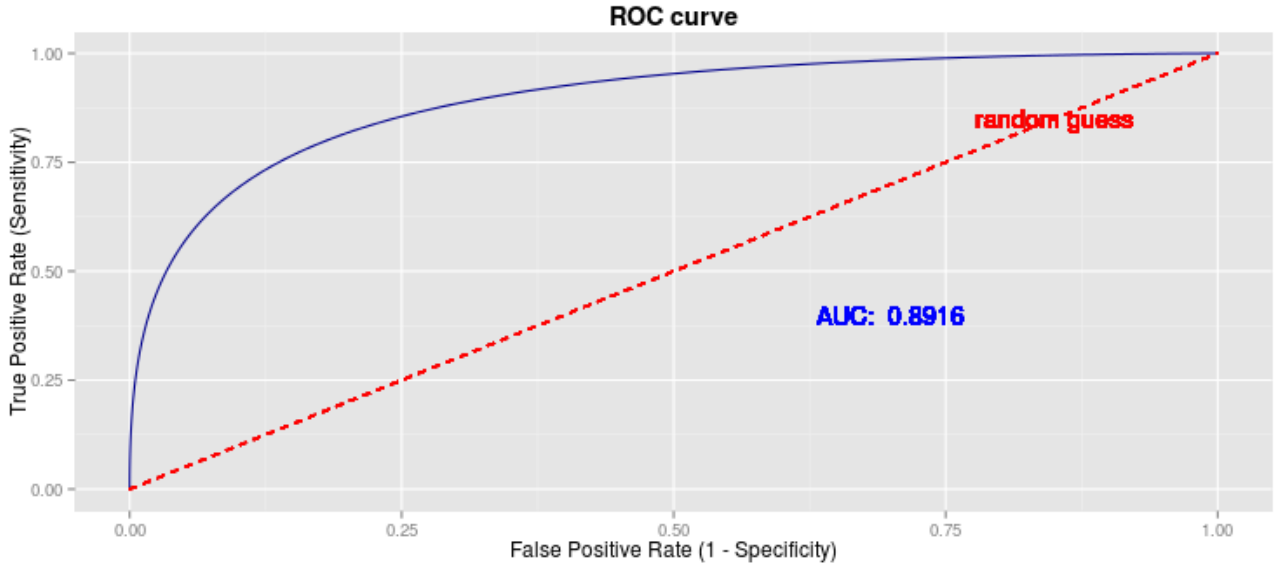
$$TPR(c) = P[Y \geq c|D = 1]$$
$$\text{versus } FPR(c) = P[Y \geq c|D = 0]$$

  - A test that can't discriminate between cases and non-cases produces a 45 degree line



- A good discriminator produces a curve that concentrates in the upper left corner





https://kennis-research.shinyapps.io/ROC-Curves/

- “Optimal” cut-off point depends on the costs associated with a false positive and a false negative
  → Is one worse than the other?
  → Commonly assumed to have equal cost
  → In some cases it’s more important to have a higher specificity (or vice versa)

- Parametric and nonparametric ROC approaches

- AUC of the ROC is the probability that a randomly selected case will score higher than a randomly selected non-case (Hanley and McNeill, 1982)
  → standard errors can be estimated via bootstrap sampling

Accommodating covariates in ROC analysis

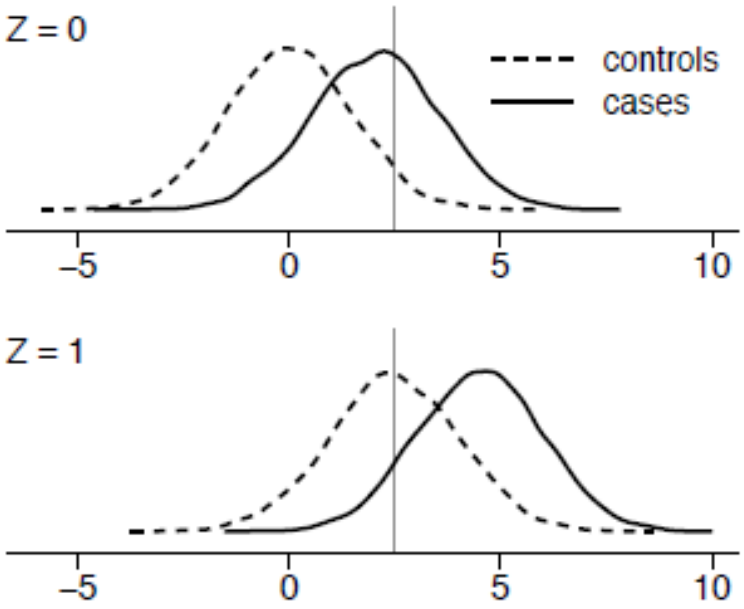
- Other factors beyond the condition status and diagnostic test (EPDS) may affect ROC analysis.

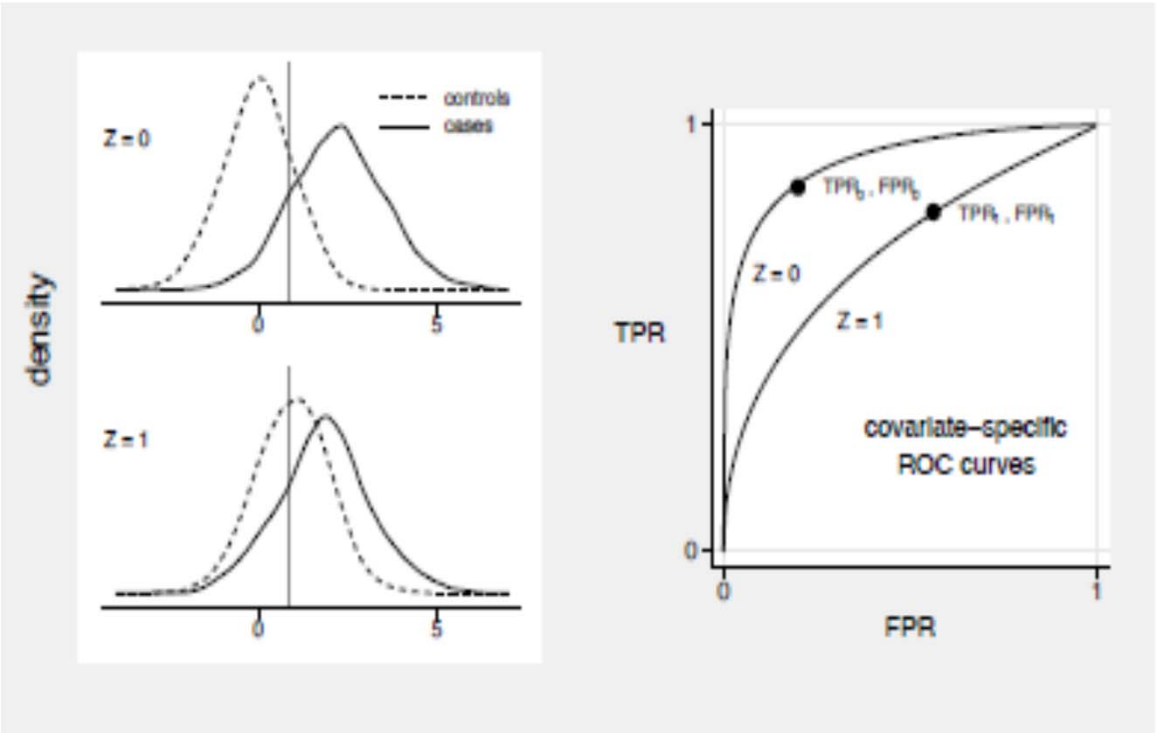
- Some covariates may impact diagnostic test distribution among controls.

  E.g., test centre may affect the control distribution of the diagnostic test

- Some covariates could affect the inherent discriminatory accuracy of the measure (i.e., the ROC curve)

  E.g., Disease severity could affect the distribution of the standardised diagnostic test under the case population - less severe cases more difficult to distinguish from controls





Figures 1 and 4 (edited) from Janes, Longton and Pepe (2009).
LHS: Covariate adjustment in controls.
RHS: Modelling ROC as a function of covariates

Pepe et al. approach

Pepe (2000); Alonzo and Pepe (2002); Janes, Longton and Pepe (2009) discuss how covariates can be incorporated into the ROC analysis to increase accuracy

1. Covariate adjustment for factors that affect marker observations among controls
2. **Model the ROC curve as a function of covariates that inherently affect discriminatory accuracy of diagnostic test**
3. Combine diagnostic test information with covariates that are known to contribute to discrimination (e.g., baseline risk factors) and determine how much discriminatory accuracy the diagnostic test adds to the known classifiers

- Pepe (2000, 2003) describe how ROC analysis can be performed as a two-stage process
  Stage 1: The control distribution of the classifier (EPDS) is estimated, assuming a normal model or using a distribution-free estimation technique. The specificity = percentiles of classifier values in control population. The classifier is standardised using the control distribution to 1 – percentile value, the false-positive rate.
  Stage 2: The ROC curve is modelled as the cumulative distribution of the case population's "false-positive" rates → sensitivity as a function of false-positives

- Covariates may enter the model at stage 1 or 2 of estimation.
  →Can use the same or different covariates at each stage
- Can't incorporate covariates into nonparametric ROC curve (stage 2) analysis in existing procedures (can only adjust control distribution/stage 1)

EPDS ROC curve with covariate adjustment

- **Here we are assuming that the covariates affect the ROC curve only**
- The associated Stata command is **rocreg** (using the *roccov* option), which models the ROC curve as a function of covariates

Will adjust the ROC curve for each of the following variables (one at a time):

- Race
- Income
- Age
- Whether they needed an interpreter
- Country of origin
- Education level

Results

Demographics of study sample N = 545

Variable	Summary
Age	Mean: 32.79, range: 16-47.5
Ethnicity	White: 284 (51.7%) Black: 177 (32.5%) Asian: 25 (4.6%) Other: 59 (10.8%)
Country of birth	UK: 262 (48.1%) Europe: 74 (13.6%) Africa: 116 (21.3%) Asia/Australasia: 38 (7%) North America/Central America/Caribbean: 27 (5%) South America: 28 (5.1%)

Demographics of study sample cont.

Need an interpreter	Yes: 40 (7.3%) No: 505 (92.7%)
Income	< £15000: 76 (13.9%) £15,000-£30,999: 71 (13%) £31,000-£45,999: 60 (11%) £46,000-£60,999: 63 (11.6%) £61,000 or more: 145 (26.6%) Would rather not say: 124 (22.8%)
Education	No formal qualifications/High school: 120 (22%) Relevant professional training/ Higher national Certificate/diploma: 142 (26.1%) Bachelor's degree: 190 (34.9%) Postgraduate: 103 (18.9%)

Observed and weighted frequencies of depression for Whooley tve and nve

Observed frequencies N = 545

	Whooley nve	Whooley tve
No depression	241	158
SCID depression diagnosis	17	129

depressed = major depressive disorder/episode, mixed anxiety/depression

Apply expansion weights

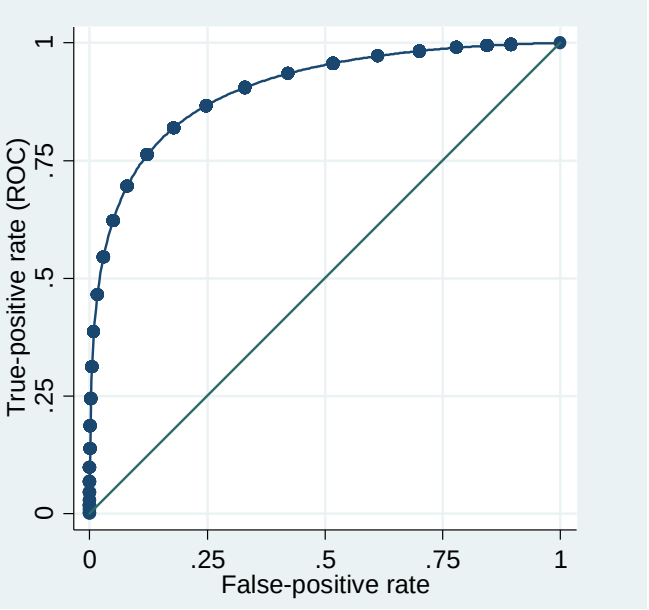
906/287=3.16 for Whooley tve and 9057/258 = 35.1 for Whooley nve

Estimated frequencies in screened sample

	Whooley nve	Whooley tve
No depression	8460 (93.4%)	499 (55.3%)
SCID depression diagnosis	597 (6.6%)	404 (44.7%)

Sensitivity, specificity, PPV and NPV of Whooley Questionnaire

Weighted	Sensitivity	Specificity	PPV	NPV
No	88%	60%	45%	93%
Yes	40.6%	94.4%	45%	93%



Provisional EPDS Results (weighted)

Cut-off	Sensitivity	Specificity	PPV	NPV
11	0.73	0.88	0.41	0.97
12	0.67	0.92	0.50	0.96
13	**0.59**	**0.95**	**0.57**	**0.95**
14	0.46	0.96	0.57	0.94

Observed and weighted frequencies of depression for EPDS (using cut-off of 13)

Observed frequencies N = 545

	EPDS <=12	EPDS>=13
No depression	351	48
SCID depression diagnosis	48	98

Apply expansion weights
906/287=3.16 for Whooley tve and 9057/258 = 35.1 for Whooley nve

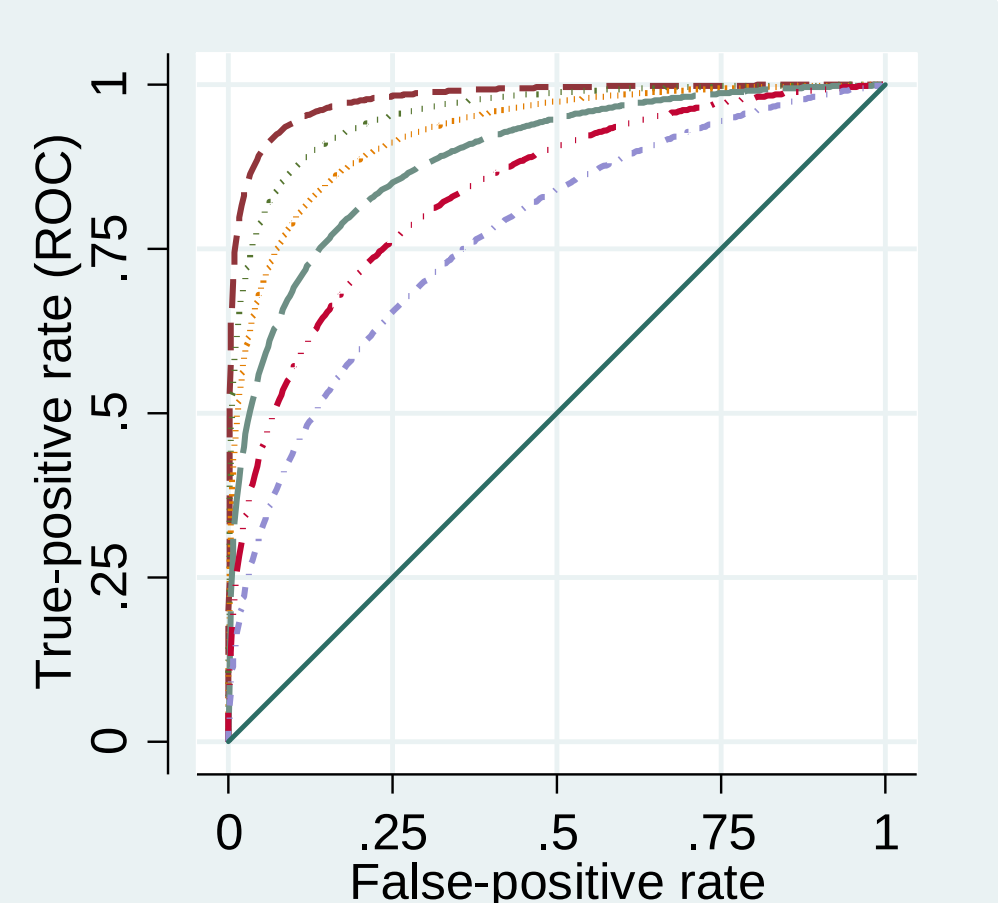
Estimated frequencies in screened sample

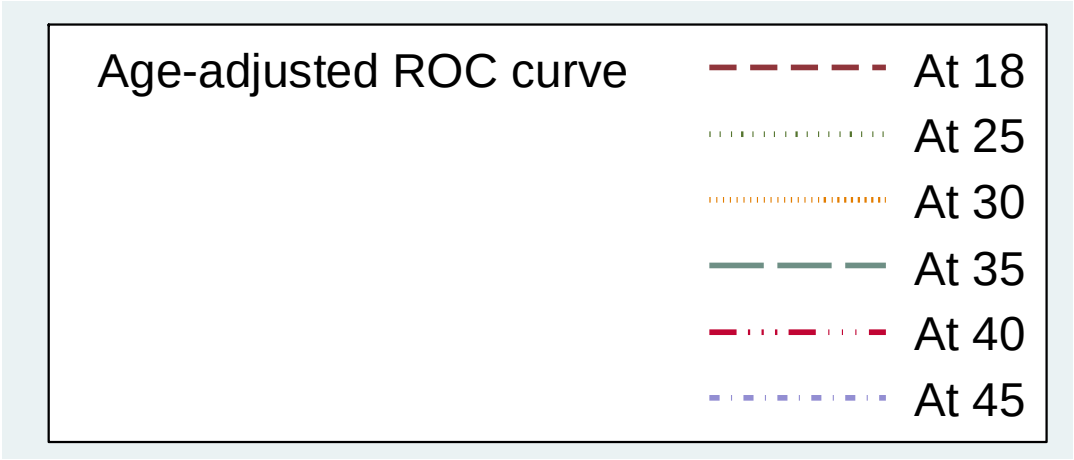
	EPDS <=12	EPDS>=13
No depression	8264 (95.3%)	695 (53.9%)
SCID depression diagnosis	407 (4.7%)	594 (46.1%)

Covariate-adjusted ROC curves

Variable	Chi-sq	Df	P-value
Race	2.73	3	0.4352
Age (continuous)	18.66	1	**<0.0001**
Income	6.37	4	0.1732
Interpreter	5.01	1	**0.0251**
Country of birth	13.84	5	**0.0166**
Qualifications	3.62	3	0.3054

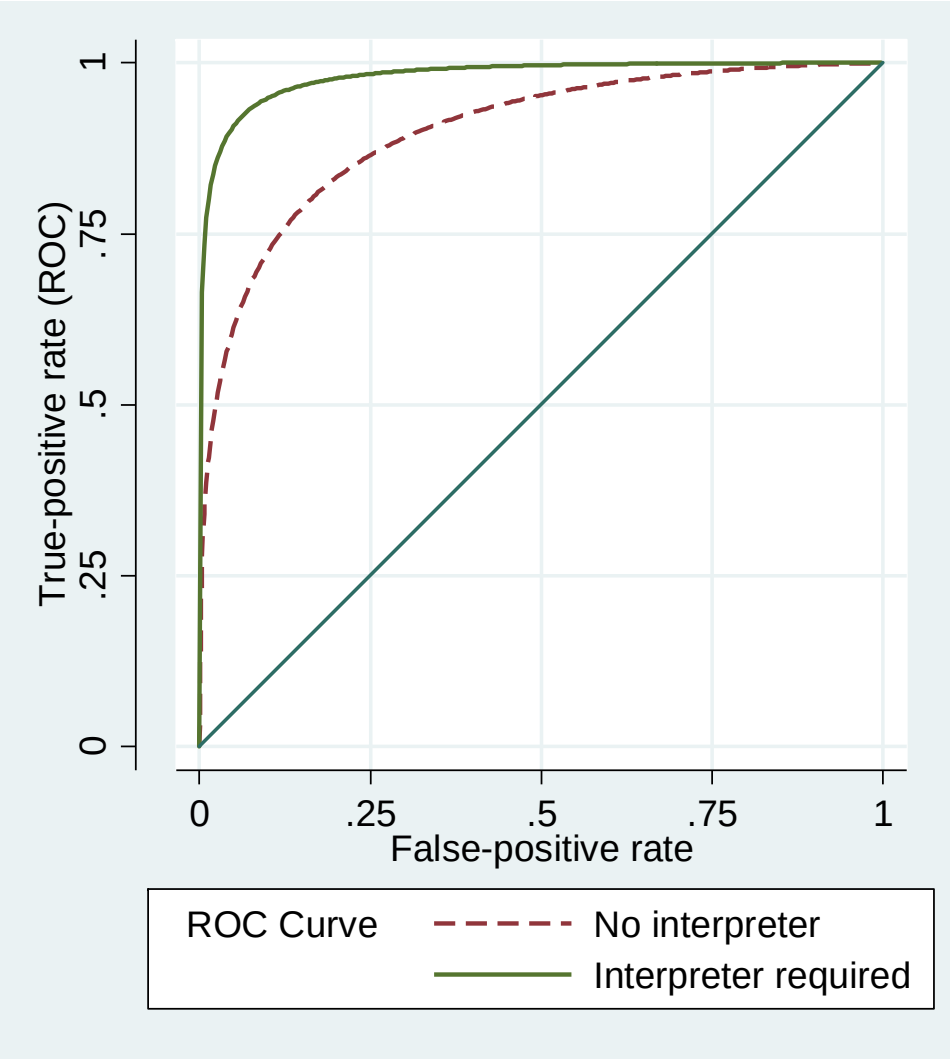
Age-adjusted ROC curves





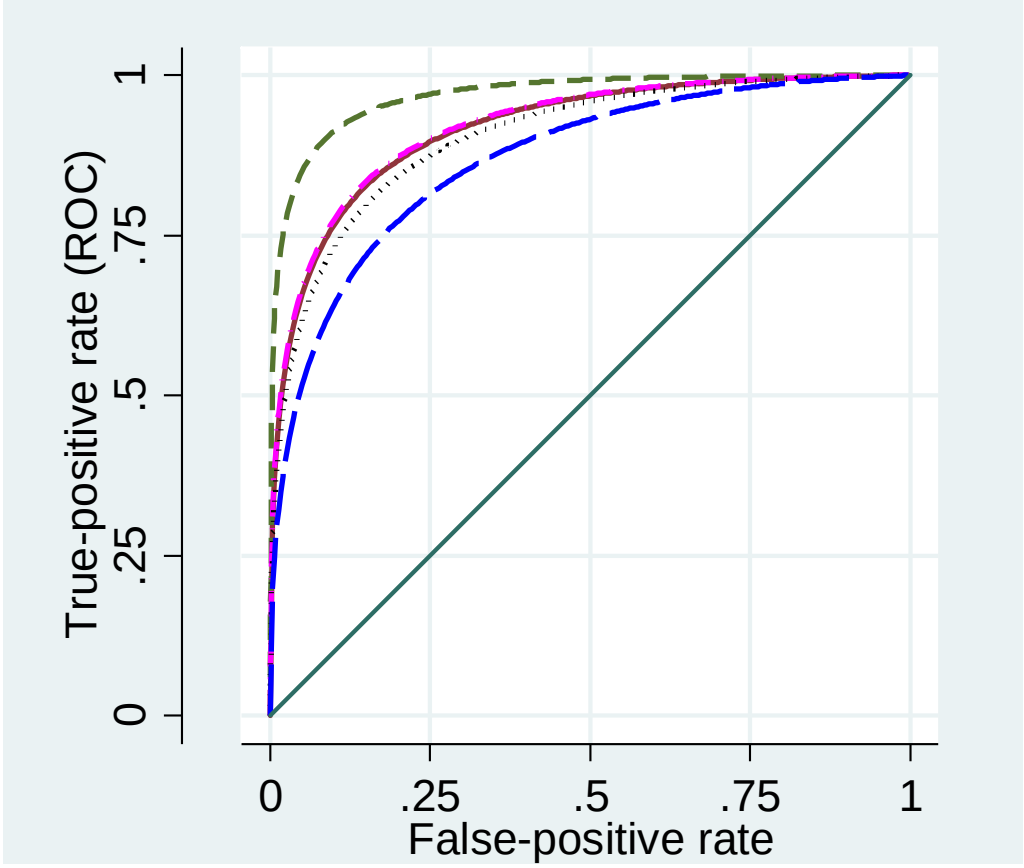
Discriminatory ability of EPDS appears to decrease with age

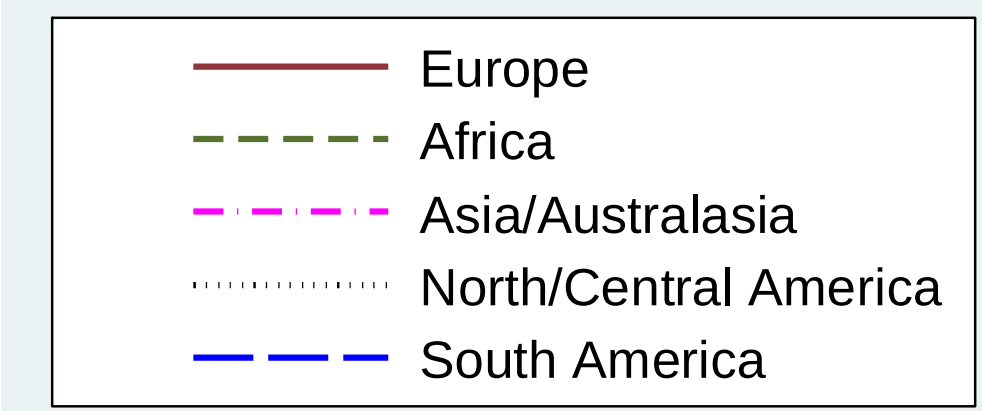
Interpreter-Adjusted ROC curve



Discriminatory ability of EPDS appears to be better in those that required an interpreter
- only 40/545 patients

Adjusted for Country of Birth





Africa found to be significantly different (p=0.005) and have best discrimination between cases and non-cases

UK was the reference category

Comparison of Whooley and EPDS (Provisional results)

Measure	Sensitivity	Specificity	PPV	NPV	AUC for ROC curve
Whooley	40.6%	94.4%	45%	93%	0.37
EPDS (cut off 13)	59%	95%	57%	95%	0.89

Discussion/Conclusions

• Whooley questionnaire had low sensitivity, but high specificity in our sample
- very important not to diagnose someone as being “not depressed” when in fact they do have depression
- false positives not such a problem as further interviews/questionnaires would reveal patient does not have depression

• Once sampling weights were accounted for, EPDS generally outperformed Whooley questionnaire

• Current data suggests a cut off value of 13 for EPDS at antenatal booking, although further investigation is needed

• Still collecting follow up data

• Some evidence that discriminatory accuracy of EPDS may differ between ages, whether an interpreter was required or not, and country of birth

This presentation summarises independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research (PGfAR) Programme (Grant Reference Number: RP-PG-1210-12002) and the National Institute for Health Research (NIHR) / Wellcome Trust King's Clinical Research Facility. The study team acknowledges the study delivery support given by the South London NIHR Clinical Research Network. Louise Howard is also supported by an NIHR Research Professorship (NIHR-RP-R3-12-011).

This study represents independent research supported by the and the NIHR Biomedical Research Centre and Dementia Unit at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

References

1. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol. 2005;106(5):1071-83.
2. Arroll B, Khin N, Kerse N. Two verbally asked questions are simple and valid. Brit Med J. 2003;327:1144-6.
3. Cox JL, Holden JM, Sagovsky R. Detection of Postnatal Depression - Development of the 10-Item Edinburgh Postnatal Depression Scale. Br J Psychiatry. 1987;150:782-6.
4. Gibson J, McKenzie-Mcharg K, Shakespeare J, Price J, Gray R. A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. Acta Psychiatr Scand. 2009;119(5):350-64.
5. First MB, Spitzer RL, Gibbon M, Williams JB. Structured Clinical Interview for DSM-IV® Axis I Disorders (SCID-I), Clinician Version, Administration Booklet: American Psychiatric Pub; 2012.
6. American Psychiatric Association. Diagnostic and Statistical Manual-Text Revision (DSM-IV-TRim, 2000): American Psychiatric Association; 2000.
7. Hanley, JA and McNeill, BJ. The meaning and use of the area under a receiver operating characteristic curve. Radiology. 1982. 143:29-36
8. Pepe, MS. An interpretation for the ROC curve and inference using GLM procedures. Biometrics. 2000. 56: 352–359.
9. Pepe, MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press; United Kingdom: 2003.
10. Alonzo, TA, and M. S. Pepe. Distribution-free ROC analysis using binary regression techniques. Biostatistics. 2002. 3: 421–432.
11. Janes, H, Longton, G, and Pepe, MS. Accommodating covariates in receiver operating characteristic analyses. The Stata Journal. 2009. 9(1): 17-39

Comparison of Whooley and EPDS (Provisional results)

Analysis	Measure	Weighted	Sensitivity	Specificity	PPV	NPV	AUC for ROC curve
1	Whooley	No	88%	60%	45%	93%	0.53
1	Whooley	Yes	40.6%	94.4%	45%	93%	0.37
2	Whooley	No	87.7%	64.5%	54.7%	91.5%	0.56
2	Whooley	Yes	39.1%	95.3%	54.7%	91.5%	0.36
1	EPDS (13)	No	66.7%	90.2%	71.6%	88%	0.87
1	EPDS (13)	Yes	59%	95%	57%	95%	0.89
2	EPDS (13)	No	61.6%	93%	81.3%	82%	0.87
2	EPDS (13)	Yes	52.6%	95.7%	64.4%	93.1%	0.87