

Validity and utility of instruments for screening of depression in women attending antenatal clinics in Blantyre district in Malawi

G Chorwe-Sungani^{a,b,*} and J Chipps^{a,b} 

^aKamuzu College of Nursing, University of Malawi, Blantyre, Malawi

^bUniversity of the Western Cape, Bellville, South Africa

*Corresponding author, email: genesischorwe@kcn.unima.mw



Introduction: Screening instruments should be brief, valid and easy to use if they are to be useful in a busy antenatal clinic in low-resource settings. A short instrument can be used in a busy antenatal clinic in combination with a more detailed instrument once referred. This study aimed at assessing the validity of a range of depression screening instruments and to test the utility of combining these instruments for use in antenatal clinics in Blantyre district, Malawi.

Methods: This was a sensitivity analysis study using a sub-sample of 97 pregnant women drawn from a cross-sectional study (sample size = 480) that was screening for depression in eight antenatal clinics. Data from the cross-sectional study for the 97 pregnant women on the 3-item screener, Edinburgh Postnatal Depression Scale (EPDS), Hopkins Symptoms Checklist-15 (HSCL-15) and Self-Reporting Questionnaire (SRQ), was compared with a gold standard, the Mini International Neuropsychiatric Interview (MINI). Sensitivity, specificity and area under curve (AUC) were calculated to test for validity of the instruments. The utility of various combinations of the instruments was tested using the compensatory, conjunctive, probability and sequential rules.

Results: The 3-item screener, EPDS, HSCL-15 and SRQ were valid instruments for screening antenatal depression. Sequential combination of the 3-item screener and SRQ had superior discriminant ability over similar combinations of the 3-item screener and either EPDS or HSCL-15 (sensitivity = 78%, specificity = 88%, AUC = 0.885).

Discussion: The 3-item screener, EPDS, HSCL-15 and SRQ are valid instruments for screening depression in local antenatal clinics. The sequential combination of the 3-item screener and SRQ may be a practical, accurate and suitable method for multistage screening of depression in antenatal clinics in Blantyre district, Malawi.

Keywords: Antenatal, depression, screening instrument, utility, validity

Introduction

Depression is a mood disorder largely characterised by low mood and lack of interest or pleasure,¹ which can affect women during pregnancy. In sub-Saharan Africa, prevalence of antenatal depression ranges from 21% to 47%, significantly contributing to the disease burden of women.^{2,3} Depression may cause fatigue, poor concentration and feelings of hopelessness in a pregnant woman.⁴ It is often associated with premature birth, intrauterine growth restriction and low birthweight.⁵ However, depression is often under-diagnosed by treating health professionals,⁶ especially in antenatal care as is seen in Malawi. In that country, midwives generally focus on the physical health of pregnant women and their babies at the expense of mental health.

Pregnant women with depression can be identified through routine screening in antenatal clinics.⁷ An instrument for screening of depression should be accurate, reliable and valid to use in antenatal clinics. Screening instruments cannot be valid without being reliable.⁸ A reliable instrument for screening of depression should be able to measure depression in pregnant women consistently.⁸ According to Wong and Lim, a valid instrument should have an ability to measure what it is supposed to measure.⁹ This is determined by its sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV).⁹ PPV and NPV measure the likelihood that a positive or negative screening test result is accurate for an individual.¹⁰ An instrument with high specificity and PPV 'rules IN' the disease while the one with high sensitivity and NPV 'rules OUT' the disease.¹¹ Sensitivity and specificity of a screening instrument are often in balance and can vary depending on cut-off scores.

Optimum cut-off scores are recommended through using a Youden index.¹²

For effective depression screening in antenatal clinics in low-resource settings, instruments should be accurate. Accuracy refers to the degree to which a measurement represents the true value of an attribute being measured.¹³ This can be determined by comparing results from a screening instrument with results generated by a gold standard using scores for area under curve (AUC),¹³ sensitivity and specificity.¹⁴ In this context, the terms accuracy and validity can be used synonymously. Screening instruments that are validated in specific settings such as antenatal clinics have a high likelihood of generating accurate results¹⁵ and may reduce under-diagnosis of depression in these settings. However, screening instruments are not a replacement for gold-standard diagnostic assessments for depression, such as the Mini International Neuropsychiatric Interview (MINI).¹⁶

Lastly, to be effective in a busy antenatal setting, screening instruments should be brief and easy to use.¹⁷ The literature suggests that brief screening instruments have greater utility in low-resource settings.¹⁸ There are reports which show that Edinburgh Postnatal Depression Scale (EPDS) and Self-Reporting Questionnaire (SRQ) have been used in research to detect antenatal depression in Malawi.² For these instruments to be considered suitable for use in low-resource settings, they should be easy to administer and acceptable for use by midwives in busy and usually understaffed antenatal clinics.⁷ Sometimes brief screening instruments may be considered as too long and time consuming for routine screening,¹⁹ especially in low-

resource settings. As such, the use of ultra-brief screening instruments which have a maximum of four items or fewer and requiring less than 2 min to administer can be suitable when using staged screening²⁰ for depression in antenatal clinics with increased workloads.¹⁷

Screening in stages may involve a two-step process where a short screening instrument is used to identify potential cases.²¹ For those who screen positive (cases), a second, often more detailed instrument with greater specificity is used to confirm caseness.²¹ This approach may be appropriate in busy antenatal settings which are not directly tasked to screen for depression as a key task. As such, the use of an ultra-brief screening instrument as the first step in screening in combination with a brief screening instrument (to be completed on a smaller group of initial screen positives) may be recommended in these settings. Screening instruments can be combined using compensatory, conjunctive, probability and sequential rules.²² It is important that if screening instruments or a combination of instruments are considered for screening in antenatal settings, these should be reliable and valid in detecting individuals²³ with depression in this setting. A study was conducted to assess the validity of a range of instruments for screening of depression and to test the utility of combining these instruments for use in antenatal clinics in Blantyre district, Malawi.

Materials and methods

This was a sensitivity analysis study, which used a sub-sample drawn from a cross-sectional study (sample size = 480) that was screening for depression using the 3-item screener, EPDS, Hopkins Symptoms Checklist-15 (HSCL-15) and SRQ in eight antenatal clinics in Blantyre district from January to May 2016. A sample size for this sensitivity analysis study was calculated using a sample size calculator.²⁴ It was estimated that the prevalence of depression among pregnant women in Malawi is 21%.² Using 95% significance level, 7.12% confidence interval, proportion of 21% and 480 (sample size for cross-sectional study) as population, a sub-sample of 100 was calculated to be sufficient for this study. A research assistant randomly selected a sub-sample of 100 pregnant women who were participating in a cross-sectional study that was going on in the eight antenatal clinics, to be interviewed further by the researcher using the MINI. The research assistant sent every third pregnant woman for further interview using the MINI, after randomly picking the first one until the desired sub-sample for each of the eight antenatal clinics was achieved. Three pregnant women declined resulting in a sub-sample of 97 pregnant women (Ndirande [$n = 25$], Limbe [$n = 23$], Mdeka [$n = 14$], Zingwangwa [$n = 10$], Chilomoni [$n = 8$], Mpemba [$n = 7$], Chileka [$n = 6$] and Lirangwe [$n = 4$] health centres) participating in this sensitivity analysis study. The inclusion criterion for this study was accepting to undergo a further interview on the same day after participating in the cross-sectional study and those who declined were excluded.

Screening instruments

This study used HSCL-15, SRQ and EPDS because they were identified as effective screening instruments for antenatal depression in low-resource settings.²⁵ The 3-item screener for depression was included because it has been recommended that valid ultra-brief instruments for screening of depression may be more suitable in detecting possible cases of depression in primary care.^{26,27} The MINI was also used because it was identified as the most widely used gold standard in low-resource settings.²⁵

The 3-item screener consisted of two ultra-brief depression screening instruments—Whooley's questions²⁸ and the one-item screening question.⁶ The Whooley's questions screen for sadness and loss of interest in the past month. The maximum total score for the 3-item screener was 3 and cut-off was set as ≥ 1 because each of the two instruments comprising the 3-item screener have a cut-off = 1. Unlike the 3-item screener, the HSCL-15 consists of 15 items of HSCL-25, a self-report inventory, which assesses for depressive symptoms a person has been bothered by in the past seven days.²⁹ Each item is rated on a Likert scale of 1–4 and the average of the 15 items is the depression score at a cut-off ≥ 1.75 . Maximum average score for HSCL-15 is 4.

With regard to the SRQ, it was designed for screening psychiatric symptoms experienced by an individual in the previous four weeks and consists of 20 questions.³⁰ The instrument has a maximum total score of 20 with a standard cut-off ≥ 10 .³¹ As for the EPDS, it is a 10-item self-reported questionnaire which measures depressive symptoms experienced in the past seven days and each item is rated on four exclusive scores (0–3).³² The instrument has a maximum total score of 30 with a standard cut-off ≥ 10 .³³ As a gold standard, the MINI is a brief structured diagnostic interview for the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV),¹⁶ which was used to confirm presence or absence of depression in pregnant women in this study.

Translation of instruments

Previously validated Chichewa-language versions of EPDS and SRQ were used in this study.³⁴ The HSCL-15, the MINI and the 3-item screener were translated into Chichewa by the first author and a social worker based on the minimum standards (back-translation and monolingual testing) for applying an instrument that was developed in another language.³⁵

Data collection

This study used data from a sub-sample of respondents ($n = 97$) who participated in a cross-sectional study that was screening for depression using the 3-item screener, EPDS, HSCL-15 and SRQ. The research assistant (registered midwife) trained in administration of data-collection instruments collected data for the cross-sectional study. In addition, he recruited a sub-sample ($n = 97$) of respondents from the cross-sectional study for further interview using the MINI in this sensitivity analysis study. The first author, a mental health nurse, administered the MINI to all respondents who agreed to participate in the sensitivity analysis study to confirm the presence or absence of depression in respondents on the same day. The first author was blind to the respondents' initial screening outcomes in the cross-sectional study. Due to the low literacy levels, the interviewer read the questions and recorded the answers on behalf of respondents.

Data analysis

Data were analysed using Statistical Package for Social Sciences (SPSS®) version 22.0 (IBM Corp, Armonk, NY, USA) and MedCalc® (www.medcalc.org). Prior to data analysis, respondents' outcomes on the MINI were extracted and entered into IBM SPSS® 22.0 together with their data from the cross-sectional study for EPDS, HSCL-15, SRQ and the 3-item screener. Prevalence of depression as determined by the MINI was calculated. A chi-square test was used to test for significant differences between demographic characteristics and depression prevalence. The reliability of each screening instrument was calculated using Cronbach's alpha. In testing for validity of these instruments,

Bayesian 2 x 2 tables and the MINI diagnosis of depression as the gold standard were used to compute sensitivity and specificity. PPV and NPV were also calculated to determine the predictive ability of the screening instruments. Receiver operating characteristics (ROC) curve analysis was used to generate AUC, standard cut-off scores, and Youden indices with their associated sensitivity and specificity for each instrument. Utility of combinations of the 3-item screener with either EPDS or HSCL-15 or SRQ to detect depression were tested using compensatory ('OR') rule, conjunctive ('AND') rule, probability rule and sequential rule. Odds ratios were computed to test the ability of individual instruments and combinations of instruments to predict antenatal depression.

Findings

A total of 97 pregnant women agreed to participate in the sensitivity analysis study. The respondents were from rural (32%, $n = 31$) and urban (68%, $n = 66$) areas of Blantyre district. More than half of them (53.6%, $n = 52$) had secondary education or above, most were married (74.2%, $n = 72$) and more than two-

thirds were unemployed (71.1%, $n = 69$). The prevalence of major depression based on the MINI in this sample was 25.8% ($n = 25$). Major depression was most prevalent amongst unmarried (88%, $n = 22$) and unemployed pregnant women (80%, $n = 20$). Age (mean = 26 ± 5.7 years), number of pregnancies (mean = 2.5 ± 1.4) and gestation periods (mean = 27.9 ± 8.1 weeks) for respondents with depression were comparable to those without depression (Table 1).

Validity of screening instruments

The 3-item screener (cut-off ≥ 1), HSCL-15 (cut-off > 1.75), SRQ (cut-off ≥ 10) and EPDS (cut-off ≥ 10) were all valid when standard cut-off scores as specified by the developers of the tools were applied (sensitivity = 60–80%, specificity = 81–97%, PPV = 59–88%, NPV = 88–92%). The 3-item screener, HSCL-15, SRQ and EPDS levels of accuracy (AUC) were ≥ 0.85 .

The 3-item screener at cut-off ≥ 1 was found to be a reliable (Cronbach's alpha = 0.7), accurate (AUC = 0.85) and valid instrument for screening depression among pregnant women. It

Table 1: Relationship between demographic characteristics of respondents and depression

Item	Depression 25 (25.8)	No depression 72 (74.2)	Total 97 (100)	Chi-square statistic	<i>p</i> -value
Occupation				1.4	0.53
Unemployed	20 (80)	49 (68.1)	69 (71.1)		
Employed	2 (8)	7 (9.7)	9 (9.3)		
Small-scale business	3 (12)	16 (22.2)	19 (19.6)		
Education level				1.3 0	0.26
Primary school or none	14 (56)	31 (43.1)	45 (46.4)		
Secondary school or above	11 (44)	41 (56.9)	52 (53.6)		
Marital status				1.9	0.33
Married	3 (12)	69 (95.8)	72 (74.2)		
Unmarried	22 (88)	3 (4.2)	25 (25.8)		
Setting				1	0.32
Urban	15 (60)	51 (70.8)	66 (68)		
Rural	10 (40)	21 (29.2)	31 (32)		
Age in years	26 \pm 5.7	25.8 \pm 5.1	25.8 \pm 5.2	2.7	0.45
Gestation period in weeks	27 \pm 7.4	27.9 \pm 8.1	27.7 \pm 7.9	1.8	0.62
Number of pregnancies	2.5 \pm 1.4	2.4 \pm 1.3	2.5 \pm 1.3	4.7	0.19

Note: Data = n (%) or mean \pm standard deviation, MINI = Mini International Neuropsychiatric Interview.

Table 2: Validity of screening instruments

Instrument cut-off	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	AUC (95% CI), <i>p</i> -value	Optimum cut-off (Se, Sp, J)	Cut-off @ Se 80% Sp in %	Cut-off Se in % @ Sp 80%
EPDS ≥ 10	68 (47–85)	88 (78–94)	65 (44–83)	89 (79–95)	0.850 (0.763–0.915), $< 0.001^*$	> 6 (88%, 74%, 0.62)	$> 7, 80, 81$	$> 7, 81, 80$
HSCL-15 ≥ 1.75	72 (51–88)	93 (85–98)	78 (56–93)	91 (82–97)	0.910 (0.835–0.959), $< 0.001^*$	> 1.7 (72%, 93%, .65)	$> 1.5, 80, 82$	$> 1.5, 82, 80$
SRQ ≥ 10	60 (39–79)	97 (90–99)	88 (64–99)	88 (78–94)	0.912 (0.837–0.960), $< 0.001^*$	> 9 (72%, 96%, .68)	$> 6, 80, 83$	$> 6, 83, 80$
3-item screener ≥ 1	80 (59–93)	81 (70–89)	59 (41–75)	92 (82–97)	0.854 (0.768–0.918), $< 0.001^*$	> 1 (80%, 81%, .61)	$> 1, 80, 81$	$> 1, 80, 80$

Notes: Se = sensitivity, Sp = specificity, AUC = area under curve, CI = confidence interval, HSCL-15 = Hopkins Symptoms Checklist-15, EPDS = Edinburgh Postnatal Depression Scale, SRQ = Self Reporting Questionnaire, * = significance set at ≤ 0.05 , J = Youden index, PPV = positive predictive value, NPV = negative predictive value.

had a good balance of sensitivity = 80%, and specificity = 81%, and NPV = 92%, suggesting that it would be good for 'ruling out' depression. The optimum cut-off score of the instrument was > 1 (Youden index = 0.61) (Table 2). This demonstrated the potential of the 3-item screener as a valid ultra-brief screening instrument for depression during pregnancy. The 3-item screener was also good at predicting depression in pregnant women (OR = 4.1 [2.3–7.4], $p < 0.001$) with screen positives being four times more likely to have depression.

This study also found that HSCL-15 (cut-off ≥ 1.75) is a reliable (Cronbach's alpha = 0.85), accurate (AUC = 0.91) and valid (sensitivity = 72%, specificity = 93%) instrument for measuring depression (see Table 2). The high specificity (93%) and PPV (78%) showed that HSCL-15 could be a good instrument for 'ruling in' depression. The HSCL-15 had the second highest accuracy (AUC = 0.91) in detecting probable depression cases, confirming its utility as a screening instrument for antenatal depression. When the cut-off score was adjusted from ≥ 1.75 to > 1.7 in order to optimise sensitivity and specificity (Youden index = 0.65), the sensitivity = 72% and specificity = 93% of HSCL-15 remained constant. The HSCL-15 predicted depression in pregnant women very well (OR = 59.3 [12–123], $p < 0.001$) with screen positives being 59 times more likely to have depression.

The SRQ (cut-off ≥ 10) was found to be reliable (Cronbach's alpha = 0.86), had the highest level of accuracy (AUC = 0.912) and was a valid (sensitivity = 60%, specificity = 97%) instrument for screening depression during pregnancy (see Table 2). The instrument had high specificity (97%) and PPV (88%), confirming that it was the best instrument for 'ruling in' depression (see Table 2). The optimum cut-off score for SRQ was > 9 (sensitivity = 72%, specificity = 96%, Youden index = 0.68). SRQ predicted depression in women (OR = 1.5 [1.3–1.8], $p < 0.001$) with screen positives being twice as likely to have depression.

The EPDS (cut-off ≥ 10) was also found to be reliable (Cronbach's alpha = 0.8), accurate (AUC = 0.85) and valid (sensitivity of 68%, specificity of 88%) with a high NPV (89%) (see Table 2). The optimum cut-off for EPDS was > 6 (sensitivity = 88%, specificity = 74%, Youden index = 0.62) (see Table 2). Decreasing the cut-off score of EPDS from ≥ 10 to > 7 resulted in a good balance between sensitivity (80%) and specificity (81%). The EPDS predicted depression in pregnant women (OR = 1.2 [1.2–1.5], $p < 0.001$) with screen positives being likely to have depression.

Utility of combining depression screening instruments

The following combination rules were tested in this study: compensatory, conjunctive, probability³⁶ and sequential.²²

Compensatory ('OR') rule

The 3-item screener and either EPDS or HSCL-15 or SRQ were combined using the compensatory rule such that a respondent was considered a case if she screened positive on any of the two combined instruments. Combination of the 3-item screener and EPDS using the compensatory rule resulted in picking 49 cases, of which one case that was missed by the 3-item screener was picked up by EPDS (Table 3). The 3-item screener detected 48 cases, which included all cases identified by HSCL-15 and SRQ. There was a substantial increase in sensitivity and a drastic decrease in specificity of EPDS, HSCL-15 and SRQ when they were combined with the 3-item screener using the 'OR' rule with all combinations having sensitivity above 80% and specificity below 70%.

Table 3: Performance of individual instruments and various combinations of screening instruments

Instrument	Optimum cut-off	Se % (95% CI)	Sp % (95% CI)	AUC (95% CI), p -value
Individual test				
EPDS	> 6	88 (69–98)	74 (62–83)	0.850 (0.763–0.915), < 0.001*
HSCL-15	> 1.7	72 (51–88)	93 (85–98)	0.910 (0.835–0.959), < 0.001*
SRQ	> 9	72 (51–88)	96 (88–99)	0.912 (0.837–0.960), < 0.001*
3-item screener	> 1	80 (59–93)	81 (70–89)	0.854 (0.768–0.918), < 0.001*
Compensatory rule testing (either test is positive)				
3-item screener or EPDS ($n = 49$, 50.5%)	> 1/> 6	96 (78–100)	50 (30–70)	0.769 (0.627–0.877), < 0.001*
3-item screener or HSCL-15 ($n = 48$, 49.5%)	> 1/> 1.4	91 (72–99)	56 (35–76)	0.866 (0.737–0.947), < 0.001*
3-item screener or SRQ ($n = 48$, 49.5%)	> 1/> 6	87 (66–97)	68 (69–98)	0.885 (0.760–0.959), < 0.001*
Conjunctive rule testing (positive on both tests)				
3-item screener and EPDS ($n = 29$, 29.9%)	> 1/> 15	42 (20–67)	90 (56–100)	0.608 (0.410–0.783), 0.33
3-item screener and HSCL-15 ($n = 23$, 23.7%)	> 1/> 2.5	33 (13–59)	100 (49–100)	0.772 (0.552–0.919), 0.03*
3-item screener and SRQ ($n = 21$, 21.6%)	> 1/> 15	33 (13–59)	100 (29–100)	0.685 (0.449–0.867), 0.28
Probability combination				
3-item screener and EPDS		88 (69–97)	82 (71–90)	0.877 (0.794–0.960), < 0.001*
3-item screener and HSCL-15		88 (69–97)	88 (78–94)	0.917 (0.852–0.982), < 0.001*
3-item screener and SRQ		92 (74–99)	83 (73–91)	0.920 (0.856–0.983), < 0.001*
Sequential rule				
3-item screener → EPDS ($n = 48 \rightarrow 29$, 60.4%)	> 1/> 6	96 (78–99)	52 (31–72)	0.775 (0.631–0.883), < 0.001*
3-item screener → HSCL-15 ($n = 48 \rightarrow 23$, 47.9%)	> 1/> 1.7	78 (56–93)	80 (59–93)	0.866 (0.737–0.947), < 0.001*
3-item screener → SRQ ($n = 48 \rightarrow 21$, 43.7%)	> 1/> 9	78 (56–93)	88 (69–98)	0.885 (0.760–0.959), < 0.001*

Notes: AUC = area under curve, CI = confidence interval, HSCL-15 = Hopkins Symptoms Checklist-15, EPDS = Edinburgh Postnatal Depression Scale. SRQ = Self Reporting Questionnaire, PPV = positive predictive value, NPV = negative predictive value, Se = sensitivity, Sp = specificity, * = significance set at ≤ 0.05 .

Conjunctive ('AND') rule

Respondents who screened positive on both combined instruments were considered as cases using the conjunctive

(‘AND’) rule. All the combinations of instruments under this rule had sensitivity of $\leq 42\%$ and specificity of $\geq 90\%$ with AUCs of ≤ 0.772 (see Table 3). Furthermore, combinations of the 3-item screener and EPDS and that of the 3-item screener and SRQ under this rule were poor at discriminating probable cases from non-probable cases, $p > 0.05$.

Probability combination

Mathematical combination of screening instruments was done using logistic regression to identify combinations which had test scores that best distinguished respondents with antenatal depression from those without. All the combinations performed in this manner achieved sensitivity of $\geq 88\%$ and specificity of $\geq 82\%$ with AUCs of ≥ 0.877 (see Table 3). The probability combination of the 3-item screener and SRQ had the best level of accuracy (AUC = 0.920 [0.856–0.983]) and a good balance between sensitivity (92%) and specificity (83%). Probability combination of the 3-item screener and SRQ was the best predictor of depression (OR = 479 [49–4689], $p < 0.001$) in this study (Table 4).

Sequential rule

In sequential combination of instruments, all respondents were initially screened using the 3-item screener and all respondents who screened positive ($n = 48$) were further assessed using EPDS, HSCL-15 and SRQ. Sequential combination of the 3-item screener and other instruments increased sensitivity above that of each instrument when used alone (see Table 3). Most of the sequential combinations' validity in detecting depression decreased below that of the individual instruments. For instance, the AUC of EPDS decreased from 0.850 (0.763–0.915) to 0.775 (0.631–0.883) and specificity decreased from 81% to 52% when the 3-item screener and EPDS were sequentially combined. The sequential combination of 3-item screener (cut-off > 1) and SRQ (cut-off > 9) had a good balance between sensitivity (78%) and specificity (88%) and demonstrated superior ability in detecting depression (AUC = 0.885 [0.760–0.959]) over other sequentially combined instruments.

Discussion

Availability of an accurate and usable screening instrument helps a health-care system to use its limited resources efficiently to provide care to those who are most vulnerable.³⁷ Screening instruments with less than four questions can effectively detect depression and are considered easy to use in clinical settings.^{6,19} This is corroborated by van Heyningen *et al.*,³⁸ who asserted that a screening instrument for use in antenatal care in low-resource

Table 4: Predictive ability of probability combinations of instruments

Instrument	Wald	OR (95% CI), <i>p</i> -value	Correctly classified depression cases (%)
3-item screener and EPDS	26.5	358 (38–3 365), < 0.001*	80.4
3-item screener and HSCL-15	28.9	401 (45–3 569), < 0.001*	87.6
3-item screener and SRQ	33.9	479 (49–4 689), < 0.001*	86.6

Notes: CI = confidence interval, HSCL-15 = Hopkins Symptoms Checklist-15, EPDS = Edinburgh Postnatal Depression Scale, SRQ = Self Reporting Questionnaire, * = significance set at ≤ 0.05 .

settings should be short and quick to administer, easy to score and interpret, have good sensitivity and specificity, and should be relevant to the setting. Nonetheless, there is always a trade-off between sensitivity and specificity of any screening instrument.³⁹ A suitable screening instrument should have a minimum acceptable balance of sensitivity/specificity (80%/70%).⁴⁰ This was achieved by EPDS (sensitivity = 88%, specificity 74%, optimum cut-off > 6) and the 3-item screener (sensitivity = 80%, specificity 81%, optimum cut-off > 1), confirming their suitability for screening depression in this population. The 3-item screener had a moderate discriminant ability (AUC = 0.85) in detecting antenatal depression. The 3-item screener is advantageous over EPDS in clinical practice because it is very short, easy to administer and easy to score, making it feasible and acceptable for use in busy settings that have inadequate resources. Therefore, this study suggests that the 3-item screener may be a suitable instrument for initial depression screening in busy antenatal clinics where true and false positives would undergo further screening.

Working from the premise that midwives may be trained to screen and refer antenatal depression cases in low-resource settings,⁷ the discriminant validity of screening instruments which can complement each other in detecting a condition if they are combined⁴¹ were tested. Probability combination of the 3-item screener and SRQ provided the best discriminant ability (AUC = 0.92) in this study. Nonetheless, probability combination has limited utility in clinical practice because its outcomes scores are arbitrary and do not share attributes of either instruments combined,³⁶ making it difficult to interpret.

The most utility was achieved by sequential combination of the 3-item screener and SRQ, which had the best balance of sensitivity (78%) and specificity (88%) compared with other instruments combined at optimum cut-off scores. This suggests that a multistage process for depression screening²⁰ can be utilised to administer a combination of an ultra-brief instrument (as initial screener) followed by a more detailed instrument (only to those who initially screened positive) in busy and understaffed antenatal clinics. The 3-item screener and SRQ combination would be feasible and acceptable for use in busy local antenatal clinics where midwives may be required to participate in screening because both instruments have binary questions that would be easy to score and interpret. Screening instruments with binary questions are less time consuming, easy to score³⁸ and easily understood by illiterate pregnant women.⁴²

Implications

Screening for depression in antenatal services, which are busy and usually understaffed in low-resource settings, should be done as a multistage process²⁰ to reduce workload by referring initial screen positives only for more detailed screening. A two-step process can be used where the 3-item screener (ultra-brief instrument), would initially be used to identify potential depression cases followed by SRQ (a more detailed instrument) to confirm the cases. Referral for specialist clinical assessment will then be determined by SRQ results. It is therefore recommended that screening and referral protocols which are developed to facilitate the detection of depression during antenatal care should incorporate this two-step process for best utility and accuracy.

Limitations of this study

The limitation of this study is that it may have been affected by recall effects and response-choice order effects.⁴³

Conclusion

This study has confirmed that the 3-item screener, EPDS, HSCL-15 and SRQ are valid instruments which are effective in screening antenatal depression when applied alone. Furthermore, sequential combination of the 3-item screener and SRQ may be a possible practical, accurate and suitable method for multistage screening of antenatal depression in antenatal clinics.

Disclosure statement – The authors declare that they have no financial or personal relationship(s) which may have inappropriately influenced the writing of this article.

Acknowledgement – The authors would like to acknowledge the invaluable contribution of all who assisted with data collection and reviewing of this article. This study was funded by the University of Malawi through grant QZA-0484 NORHED 2013.

Ethics approval – This study received ethics approval from the Senate Research and Ethics Committee (University of the Western Cape) and the College of Medicine Research and Ethics Committee (University of Malawi). Pregnant women diagnosed with depression were referred to a psychiatric clinic.

ORCID

J Chipps  <http://orcid.org/0000-0002-7895-4483>

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington: American Psychiatric Publishing; 2013.
- Stewart RC, Umar E, Tomenson B, et al. A cross-sectional study of antenatal depression and associated factors in Malawi. *Arch Women's Ment Health*. 2014;17(2):145–54. <https://doi.org/10.1007/s00737-013-0387-2>
- Rochat TJ, Tomlinson M, Bärnighausen T, et al. The prevalence and clinical presentation of antenatal depression in rural South Africa. *J Aff Disord*. 2011;135(1–3):362–73. <https://doi.org/10.1016/j.jad.2011.08.011>
- Stewart RC. Maternal depression and infant growth ? a review of recent evidence. *Mater Child Nutr*. 2007;3(2):94–107. <https://doi.org/10.1111/mcn.2007.3.issue-2>
- Kinsler PA, Lyon DE. A conceptual framework of stress vulnerability, depression, and health outcomes in women: potential uses in research on complementary therapies for depression. *Brain Behav*. 2014;4(5):665–74. <https://doi.org/10.1002/brb3.2014.4.issue-5>
- Vahter L, Kreegipuu M, Talvik T, et al. One question as a screening instrument for depression in people with multiple sclerosis. *Clin Rehabil*. 2007;21(5):460–64. <https://doi.org/10.1177/0269215507074056>
- Honikman S, van Heyningen T, Field S, et al. Stepped care for maternal mental health: A case study of the perinatal mental health project in South Africa. *PLoS Med*. 2012;9(5):e1001222. <https://doi.org/10.1371/journal.pmed.1001222>
- Tavakoli M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Edu*. 2011;2:53. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Wong HB, Lim GH. Measures of diagnostic accuracy: sensitivity, specificity. *Proce Singapore Healthcare*. 2011;20(4):316–8. <https://doi.org/10.1177/201010581102000411>
- Vanderheyden AM. Technical adequacy of response to intervention decisions. *Excep Child*. 2011;77(3):335–50. <https://doi.org/10.1177/001440291107700305>
- Parikh R, Mathai A, Parikh S, et al. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45. <https://doi.org/10.4103/0301-4738.37595>
- Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32–5. [https://doi.org/10.1002/\(ISSN\)1097-0142](https://doi.org/10.1002/(ISSN)1097-0142)
- Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 2013;4(2):627.
- Henderson M. Predicting performance on high stakes testing: validity and accuracy of curriculum-based measurement of reading and writing [Doctoral Thesis]. Baton Rouge, LA: Department of Psychology, Louisiana State University; 2009.
- Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr*. 2007;96(3):338–41. <https://doi.org/10.1111/j.1651-2227.2006.00180.x>
- Sheehan DV, Lecrubier Y, Sheehan KH, et al. The mini-international neuropsychiatric interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psych*. 1998;59(20):22–33.
- Tsai AC, Scott JA, Hung KJ, et al. Reliability and validity of instruments for assessing perinatal depression in african settings: systematic review and meta-analysis. *PLoS One*. 2013;8(12):e82521. <https://doi.org/10.1371/journal.pone.0082521>
- Hanlon C, Medhin G, Selamu M, et al. Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. *J Affect Disor*. 2015;186:32–9. <https://doi.org/10.1016/j.jad.2015.07.015>
- Lombardo P, Vaucher P, Haftgoli N, et al. The 'help' question doesn't help when screening for major depression: external validation of the three-question screening test for primary care patients managed for physical complaints. *BMC Med*. 2011;9(1):1.
- Fiest KM, Patten SB, Wiebe S, et al. Validating screening tools for depression in epilepsy. *Epilepsia*. 2014;55(10):1642–50. <https://doi.org/10.1111/epi.2014.55.issue-10>
- Reme SE, Lie SA, Eriksen HR. Are 2 questions enough to screen for depression and anxiety in patients with chronic low back pain? *Spine*. 2014;39(7):E455. <https://doi.org/10.1097/BRS.0000000000000214>
- Ramlall S, Chipps J, Bhigjee A, et al. The sensitivity and specificity of subjective memory complaints and the subjective memory rating scale, deterioration cognitive observee, mini-mental state examination, six-item screener and clock drawing test in dementia screening. *Dement Geriatr Cogn Disord*. 2013;36(1–2): 119–135. <https://doi.org/10.1159/000350768>
- Zhu W, Zeng N, Wang N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. Paper presented at NESUG 2010: Health care and life sciences; 27th May 2015, 2010; Baltimore.
- Calculator.net. Sample size calculator [cited 2015 Feb 20]. 2015. Available from <http://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=7.12&pp=21&ps=480&x=45&y=13>.
- Chorwe-Sungani G, Chipps J. A systematic review of screening instruments for depression for use in antenatal services in low resource settings. *BMC Psych*. 2017;17(1):12. <https://doi.org/10.1186/s12888-017-1273-7>
- Bosanquet K, Bailey D, Gilbody S, et al. Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. *BMJ Open*. 2015;5(12):e008913. <https://doi.org/10.1136/bmjopen-2015-008913>
- Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? *Br Journal Gen Pract: J Royal Coll Gen Pract*. 2007;57(535):144–51.
- Whooley MA, Avins AL, Miranda J, et al. Case-finding instruments for depression. *J Gen Intern Med*. 1997;12(7): 439–445. <https://doi.org/10.1046/j.1525-1497.1997.00076.x>
- Derogatis LR, Lipman RS, Rickels K, et al. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci*. 1974;19(1):1–15. [https://doi.org/10.1002/\(ISSN\)1099-1743](https://doi.org/10.1002/(ISSN)1099-1743)
- Beusenbergh M, Orley JH, World Health Organization. A User's Guide to the Self Reporting Questionnaire (SRQ). Geneva: World Health Organisation; 1994.
- Kumbhar UT, Dhumale GB, Kumbhar UP. Self reporting questionnaire as a tool to diagnose psychiatric morbidity. *Natl J Med Res*. 2012;2:51–4.

32. Tran TD, Biggs B-A, Tran T, et al. Perinatal common mental disorders among women and the social and emotional development of their infants in rural Vietnam. *J Aff Disor.* 2014;160:104–12. <https://doi.org/10.1016/j.jad.2013.12.034>
33. Martins CdSR, Motta JvdS, Quevedo LA, et al. Comparison of two instruments to track depression symptoms during pregnancy in a sample of pregnant teenagers in Southern Brazil. *J Aff Diso.* 2015;177:95–100. <https://doi.org/10.1016/j.jad.2015.01.051>
34. Stewart RC, Umar E, Tomenson B, et al. Validation of screening tools for antenatal depression in Malawi—A comparison of the Edinburgh postnatal depression scale and self reporting questionnaire. *J Affec Disor.* 2013;150(3):1041–7. <https://doi.org/10.1016/j.jad.2013.05.036>
35. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs.* 2004;48(2): 175–186. <https://doi.org/10.1111/jan.2004.48.issue-2>
36. Mackinnon A, Mulligan R. Combining cognitive testing and informant report to increase accuracy in screening for dementia. *Am J Psych.* 1998;155(11):1529–35. <https://doi.org/10.1176/ajp.155.11.1529>
37. Baken DM, Woolley C. Validation of the distress thermometer, impact thermometer and combinations of these in screening for distress. *Psycho-Oncology.* 2011;20(6): 609–14. <https://doi.org/10.1002/pon.v20.6>
38. van Heyningen T, Baron E, Field S, et al. Screening for common perinatal mental disorders in low-resource, primary care, antenatal settings in South Africa. CPMH Policy Brief. Cape Town: Centre for Public Mental Health; 2014.
39. De Souza J, Jones LA, Rickards H. Validation of self-report depression rating scales in Huntington's disease. *Movement Disor.* 2010;25(1):91–6. <https://doi.org/10.1002/mds.22837>
40. Pettersson A, Boström KB, Gustavsson P, et al. Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nord J Psych.* 2015;69(7):497–508. <https://doi.org/10.3109/08039488.2015.1008568>
41. Ladeira RB, Diniz BS, Nunes PV, et al. Combining cognitive screening tests for the evaluation of mild cognitive impairment in the elderly. *Clinics.* 2009;64(10):967–73.
42. Stewart RC, Umar E, Tomenson B, et al. Validation of screening tools for antenatal depression in Malawi—A comparison of the Edinburgh Postnatal Depression Scale and Self Reporting Questionnaire. *J Affec Disor.* 2013;150(3):1041–7. <https://doi.org/10.1016/j.jad.2013.05.036>
43. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Publ Health.* 2005;27(3):281–91. <https://doi.org/10.1093/pubmed/fdi031>

Received: 13-09-2017 Accepted: 14-01-2018