Validity and accuracy of the Whooley questions to identify maternal distress in Dutch pregnant women.

Yvonne Fontein-Kuipers and Julie Jomeen

Abstract

Purpose To investigate the validity and accuracy of the Whooley questions for routine screening of maternal distress in Dutch antenatal care.

Design and methodology In this cohort design we evaluated self-reported responses to the Whooley questions against the Edinburgh Depression Scale screening for antenatal depression, State-Trait Anxiety Inventory for general anxiety and the Pregnancy Related Anxiety Questionnaire-Revised screening for pregnancy related anxiety, among Dutch pregnant women during the first and third trimester of pregnancy. We used standard diagnostic performance measures for the two case-finding items.

Findings The Whooley items in this study showed a higher specificity than sensitivity. The Whooley results showed good evidence to identify women who are depressed or (trait)anxious in both trimesters of pregnancy, but the results showed weak to moderate evidence to identify pregnancy-related anxiety. The Whooley items had a low to moderate predictive ability for depression, trait-anxiety and pregnancy-related anxiety and a good ability for negative case-finding. The Whooley items proved to be more able to report how effective the case-finding questions are in identifying women without depression, trait-anxiety and pregnancy-related anxiety (ruling out) rather than how effective these are in identifying women with depression, trait-anxiety and pregnancy-related anxiety (ruling in). The Whooley items were accurate in identifying depression and trait-anxiety in both trimesters but were not very accurate to identify pregnancy-related anxiety.

Research limitations/ implications Assessment of pregnancy-related anxiety using a case-finding tool requires further attention.

Practical implications The two-item Whooley case-finding tool has shown good utility as a screening instrument for antenatal depression and anxiety. Continuous assessment of maternal emotional health during the childbearing period or, at least, revisiting the topic, would both support the woman and the midwife in regarding perinatal emotional wellbeing as an important remit of midwifery care.

Originality/ value A novel aspect of this paper is the proposition of applying the Whooley questions at later stages of pregnancy or presenting the Whooley questions in a written or digital form.

Keywords Maternal distress; antenatal; case-finding; Whooley; antenatal depression; antenatal anxiety; pregnancy-related anxiety

Article classification Research paper

Introduction

Maternal distress refers to a spectrum of psychological symptoms during the antenatal period. Depression, stress and anxiety, either pregnancy or non-pregnancy/ birth related, are the most commonly mentioned constructs of maternal distress. There is increasing evidence that maternal distress among otherwise healthy pregnant women can be a predictor for preterm labour and low birth weight. Additionally, adverse short and long-term postpartum mental health effects have been reported for both mother and child, including post-partum depression and post-traumatic stress (Mulder *et al.*, 2002; O'Connor *et al.*, 2002; Heron *et al.*, 2004; van Son *et al.*, 2005; Leigh and Milgrom, 2008; Loomans *et al*, 2013; Blackmore *et al.*, 2016), behaviour and emotional problems in children, and problematic family relationships (Mennes *et al.*, 2006; Robinson *et al.*, 2008; Goodyer and Cooper, 2011; Glover, 2013). The pPopulation prevalence of maternal distress among Dutch women is just over 20% with varying prevalences for depression, anxiety and pregnancy-related anxiety (Fontein-Kuipers *et al.*, 2015a; Fontein-Kuipers *et al.*, 2016; Witteveen *et al.*, 2016). Dutch epidemiological research shows that antenatal depression coexists with other common emotional health symptoms such as general anxiety and pregnancy-related anxiety – known as maternal distress (Ingstrup *et al.*, 2012; Fontein-Kuipers *et al.*, 2015a; Fontein-Kuipers *et al.*, 2016). Disregarding the differences in symptoms these constructs are significantly correlated (Fontein-Kuipers *et al.*, 2015a; Fontein-Kuipers *et al.*, 2016).

Women in the Netherlands with uncomplicated pregnancies receive midwife-led care. About 80% of pregnant women start their care with midwives in primary care (Perined, 2016). Midwives have explicitly been appointed as promoters of maternal perinatal mental health (KNOV, 2010) although their responsibilities in relation to mental health promotion have not been clearly articulated. Available quality standards for midwifery primary care describe the need and the elements of screening and assessment of maternal distress, but do not further explicate this and don't provide the relevant tools in order to do so (de Boer and Zeeman, 2008; Beentjes *et al.*, 2012). Hence, it is important to find Dutch midwives a screening tool for detecting maternal distress (Fontein-Kuipers *et al.*, 2014) to offer women optimal follow-on treatment. Hence, to prevent and reduce the effects of maternal distress. Currently many women with maternal distress remain 'under the radar' and the provision of care is thus often inadequate (Ruiter *et al.*, 2010). The NICE guidelines (NICE, 2014) endorse the use of the Whooley questions to identify maternal distress, although in the Netherlands these questions are mainly utilised by General Practitioners on a general population – not specific for pregnant women (Spijker *et al.*, 2013). The Whooley questions originated from the Primary Care Evaluation of Medical Disorders Procedure (PRIME-MD) questionnaire, which has been validated for screening for depression in primary care settings (Spitzer *et al.*, 1994). More recently it has been validated within the pregnant population (Mann *et al.*, 2012; Darwin *et al.*, 2016; Howard *et al.*, 2018), acknowledging that the Whooley questions are a tool that midwives can use to assess a childbearing woman's emotional health.

As screening for maternal distress still is a rather new and underdeveloped skill for Dutch midwives (Fontein-Kuipers *et al.*, 2014), midwives might benefit from a standardized, brief, accessible, easy to administer and appropriate tool such as the Whooley case-finding questions (Bosanquet *et* al., 2015; Marsay *et al.*, 2017). Before it can be considered or recommended to implement this case-finding tool in Dutch antenatal care, the Whooley items need to be validated for its utility. We therefore investigated the

validity and the diagnostic accuracy of the Whooley questions against different constructs of maternal distress as a reference standard during pregnancy, identified by validated psychometric instruments with established cut-off points for antenatal depression, general anxiety and pregnancy-specific anxiety. We aimed to optimise current midwifery practice and increase the knowledge of midwifery practitioners – emotional health and wellbeing of childbearing women being the topic of interest.

Methods
Design and sample procedure
We conducted a cohort study including a sample of Dutch-speaking pregnant women with uncomplicated pregnancies, receiving midwife-led primary care. We included women who were pregnant with a singleton infant and who did not require obstetric-led care as a result of existing or likely complications. Women in secondary and tertiary care were excluded from our study. Recruitment and sampling procedures are reported in detail elsewhere (Fontein-Kuipers *et al.*, 2016). Self-completed questionnaires were digitally distributed to the eligible women. The data were collected between April 2013 and 6 March 2015 during the participants' first trimester of pregnancy (T1) at 7.06 (SD± 2.2, range 3–15) weeks gestation and (T2) at 37 (SD± 1.53, range 35–42) weeks gestation. The research ethics committee METC-Atrium-Orbis-Zuyd, approved the research protocol (registration no. 13-N-45 (11-N-101)/11-4-2013).

Measures
Four self-report measures were completed alongside socio-demographic information and personal details. We included Dutch versions of the Whooley questions (Spijker *et al.*, 2013), the Edinburgh Depression Scale (EDS) (Pop *et al.*, 1992), State-Trait Anxiety Inventory (STAI) (Van der Ploeg *et al.*, 1980), and the Pregnancy Related Anxiety Questionnaire-Revised (PRAQ-R2) (Huizink *et al.*, 2004; Huizink *et al.*, 2016) in our questionnaire.

Whooley questions
We used the two Whooley items (Whooley *et al.*, 1997), that identify potential low mood and loss of interest or pleasure amongst women during the perinatal period (see Box 1). Questions are answered positively (yes) or negatively (no). A subsequent third was added, the so called Arroll question - asking whether the woman wanted help with the difficulties identified in answer to one or both of two Whooley questions (Spitzer *et* al, 1994; Arroll *et al.*, 2005). A 'yes' answer was considered a positive test.

BOX 1. Whooley items
During the past month, have you often been bothered by feeling down, depressed or hopeless? (YES/ NO)

During the past month, have you often been bothered by little interest or pleasure in doing things? (YES/ NO)

Edinburgh Depression Scale (EDS)
We used the EDS, a 10-item questionnaire to screen for the likelihood of antenatal depression (Murray and Cox, 1990). We asked participants to reflect on their feelings and thoughts of the last seven days. Responses are scored 0, 1, 2 or 3 in seriousness of symptoms. The total score ranges from 0 to 30. In this study, we measured depression using a validated cut-off score of 11 or more for women in the first

trimester and 10 or more for women in the third trimester (Bergink *et al.*, 2011).

State-Trait Anxiety Inventory (STAI)
We used the Trait scale of the STAI to identify feelings of anxiety. Trait-anxiety is viewed as a more general concept of anxiety - a person's proneness to anxiety, a relatively stable personality characteristic (Spielberger *et al.*, 1970). The Trait scale contains 20 items and uses a 4-point rating scale to measure anxiety (*1 'not at all'; 4 'very'*) (Spielberger *et al.*, 1970; Nast *et al.*, 2013). Scores vary between 20 and 80. Women with scores of 41 and higher are perceived to have high levels of anxiety (Van der Ploeg *et al.*, 1980).

Pregnancy-Related Anxiety Questionnaire-Revised (PRAQ-R2)
We measured pregnancy-related anxiety with the 10-item PRAQ-R2 (Huizink *et al.*, 2004; Huizink *et al.*, 2016). The instrument measures three items: fear of giving birth, fear of bearing a physically or mentally handicapped child and maternal concern about own appearance. The PRAQ-R2 uses a 5-point rating scale to measure fear and worries (*1 'not at all'; 5 'very'*). We asked women to choose the most appropriate answer about their feelings during the current pregnancy. Scores vary between 10 and 50. We considered the 15% highest scoring women on the PRAQ-R2 total score as indicators of those with high levels of pregnancy-related anxiety (Huizink *et al.*, 2016; Witteveen *et al.*, 2016).

Data analysis
We based the sample size calculation on a previously developed method for diagnostic accuracy studies (Flahault *et al.*, 2005; Mann *et al.*, 2012) and calculated that we needed a sample size of 379 women in order to make inferences about pregnant women receiving midwife-led care from the sample. We calculated Cronbach's alpha (α) to measure internal consistency of the Whooley items and the results were considered as 'good', at > α .80 We identified the high scores of each psychometric measure and recoded these in either 'yes' when above cut-off level and 'no' when below cut-off level. Levels above cut-off points were considered a positive test. We ascertained the rates of 'true' and 'false' positives and 'true' and 'false' negatives for the Whooley questions, the EDS, STAI and PRAQ-R2. Validity was assessed using 2×2 contingency tables of weighted prevalences. Agreement between the Whooley and the EPDS, STAI and PRAQ-R2 were analysed using standard diagnostic performance measures: sensitivity (the proportion of true positives correctly identified by the test), specificity (the proportion of true negatives correctly identified by the test), likelihood ratio (shows how much more likely a woman is to get a positive test if she has maternal distress, compared with a woman without maternal distress), positive predictive value (the proportion of women with positive test results who are correctly identified) and negative predictive value (the proportion of women with negative test results who are correctly identified). Here, the EPDS, STAI and PRAQ-R2 were treated as the gold standard against which the 'test' was compared; using a positive response to either Whooley item as the criterion for possible caseness. We used the Receiver Operating Characteristics (ROC) analysis to determine the accuracy of the Whooley items, which are reported as area under the curve (AUC). AUC with a value approaching 1.0 indicated a high sensitivity and specificity (Lalkhen and McCluskey, 2008). Statistical analyses were performed using SPSS 24.0. The PRAQ-R2 subscale scores are not reported in this paper. The Arroll question is only included for descriptive analysis.

Results

Participants

We received 433 completed questionnaires at T1 and 343 at T2 (attrition 21%). As shown in Table 1, women in the full sample (T1) were aged between 18 and 42 years (Mean 30; SD 3.9). Half of the women (48.3%) were nulliparous women and half (52%) were multiparous women. A quarter of our sample (24.5%) had a history of psychological problems.

Table 1. Demographic and personal characteristics participants (n= 433)

	Mean (SD ±) range	N (%)
Age in years	30.05 (±3.9) 18-42	
Gestational age in weeks T1	7.06 (±2.2) 3-15	
Gestational age in weeks T2 (n=343)	37 (±1.5) 35-42	
Nulliparous		208 (48)
Multiparous		225 (52)
Partnered		433 (100)
Working (paid) job		205 (95.3)
Ethnicity		
Respondent born in the Netherlands		423 (97.6)
Respondent born in other Western country		5 (1.2)
Respondent born in non-Western country		5 (1.2)
Education		
Low level of education		30 (6.9)
Medium level of education		163 (37.6)
High level of education		240 (55.4)
Emotional wellbeing		
Self-reported history of psychological problems		106 (24.5)
Current medication use (prescribed) for psychological problems		14 (3.2)

Maternal Distress

The number of positive answers on one of the Whooley items and the Arroll question increased from T1 to T2 (27% to 31%). The proportion of women who met the criteria for the likelihood of depression increased from T1 to T2 (7.4% to 16.9%). The proportion of women who met the criteria for general (trait-)anxiety remained stable during the course of pregnancy (14.1% and 14.3%). For pregnancy-related anxiety, a dichotomous cut-off score of ≥37 (T1) and ≥34 (T2) was chosen, hereby identifying the 15% highest scoring women on the PRAQ-R2 total score (Witteveen *et al.*, 2016). The proportion of women who met the criteria for pregnancy-related anxiety remained stable during the course of pregnancy (15.9% and 16%). The prevalence of depression increased but general anxiety and pregnancy-related anxiety remained stable during the course of pregnancy (see Table 2).

Table 2. Population prevalences maternal distress T1 and T2

	First trimester (T1) n=433			Third trimester (T2) n=343		
	N (%)	95% CI		N (%)	95% CI	
Whooley item 1[1] *positive*	83 (19.2)			79 (23)		
Whooley item 2[2] *positive*	74 (17.1)			78 (22.7)		
Arroll [3] YES	39 (9)			106 (31)		
EDS score 11 [T1] / 10 [T2] or more	32 (7.4)	5.1 - 10.3		58 (16.9)	13.1 – 21.3	
STAI score 41 or more	61 (14.1)	11 – 17.7		49 (14.3)	10.8 – 18.5	
Top 15% PRAQ total scores	69 (15.9)	12.6 – 19.7		55 (16)	12.3 – 20.3	

[1] During the past month, have you often been bothered by feeling down, depressed, or hopeless?

[2] During the past month, have you often been bothered by little interest or pleasure in doing things?

[3] Is this something you feel you need or want help with?

Reliability Analysis Whooley Questions

At T1 reliability analysis showed α of .56, indicating a poor internal consistency of the two Whooley items. The inter-item correlation was α .39, showing a weak correlation. At T2 reliability analysis showed α of .71, demonstrating an acceptable internal consistency of the two Whooley items. The inter-item correlation was α .55, showing a modest correlation. Based on T1 we decided to analyse the Whooley items as separate measures for maternal distress, i.e. depression (EDS), state anxiety (STAI) and pregnancy-related anxiety (PRAQ-R2).

Diagnostic Accuracy of the Whooley Items for Depression

Both the Whooley items at T1 and T2 showed a higher specificity than sensitivity. During the first trimester of pregnancy, 28% and 31% women with depression would go undetected when using the Whooley items, while 15% and 13% would be incorrectly identified as being depressed. During the third trimester 26% of women would go undetected; 13% and 12% would incorrectly identified as being depressed. The likelihood ratios indicated that positive Whooley results showed good evidence to identify women who are depressed in both trimesters. The positive predictive value at T1 showed a low predictive ability of the Whooley items for depression, but this increased to modest predictive ability at T2. The negative predictive value showed a rather conclusive ability for negative case-finding results in both trimesters of pregnancy. The Whooley items showed to be better able to identify women *without* depression than those *with* depression. UAC showed that the Whooley items were accurate in identifying depression in both trimesters (see Table 3; Figure 1 to 4).

Diagnostic Accuracy of the Whooley Items for Trait-Anxiety

Both the Whooley items showed higher specificity than sensitivity at both T1 and T2. During the first trimester of pregnancy 35% and 42% of women with trait-anxiety would go undetected when using the Whooley items, while 10% and 7% would be incorrectly identified as being anxious. During the third trimester of pregnancy 20% and 31% would go undetected; 13% and 15% would be incorrectly identified as being anxious. The likelihood ratios indicated that positive Whooley results showed good evidence to identify women with trait-anxiety in both trimesters. The positive predictive value at T1 showed a modest predictive ability of the Whooley items for trait-anxiety. The negative predictive value showed a rather

conclusive ability for negative case-finding results in both trimesters. The Whooley items showed to be better able to identify women *without* trait-anxiety than those *with* trait-anxiety. UAC showed that the Whooley items were accurate in identifying trait-anxiety in both trimesters (see Table 4; Figure 1 to 4).

Diagnostic Accuracy of the Whooley Items for Pregnancy-related Anxiety
Both the Whooley items showed higher specificity than sensitivity at both T1 and T2, in specific at T1. During the first trimester of pregnancy 62% and 74% of women with pregnancy-related anxiety would go undetected when using the Whooley items, while 16% and 15% would be incorrectly identified as being anxious. During the third trimester of pregnancy 45% and 47% would go undetected; 17% would be incorrectly identified as being anxious. The likelihood ratios indicated that positive Whooley results showed weak evidence to identify women who actually have anxious feelings that are directly related to pregnancy in the first trimester of pregnancy, and modest evidence in the third trimester of pregnancy. The positive predictive value showed a low predictive ability of the Whooley items for anxiety directly related to pregnancy in both trimesters, while the negative predictive value at both T1 and T2 showed a good ability for negative case-finding results. The Whooley items showed to be better able to identify women *without* pregnancy-related anxiety than those with pregnancy-related anxiety. UAC showed that the Whooley items did not accurately identify pregnancy-related anxiety (see Table 5; Figure 1 to 4).

Figure 1. ROC analysis EDS, STAI and PRAQ-R2 at T1, Whooley 1
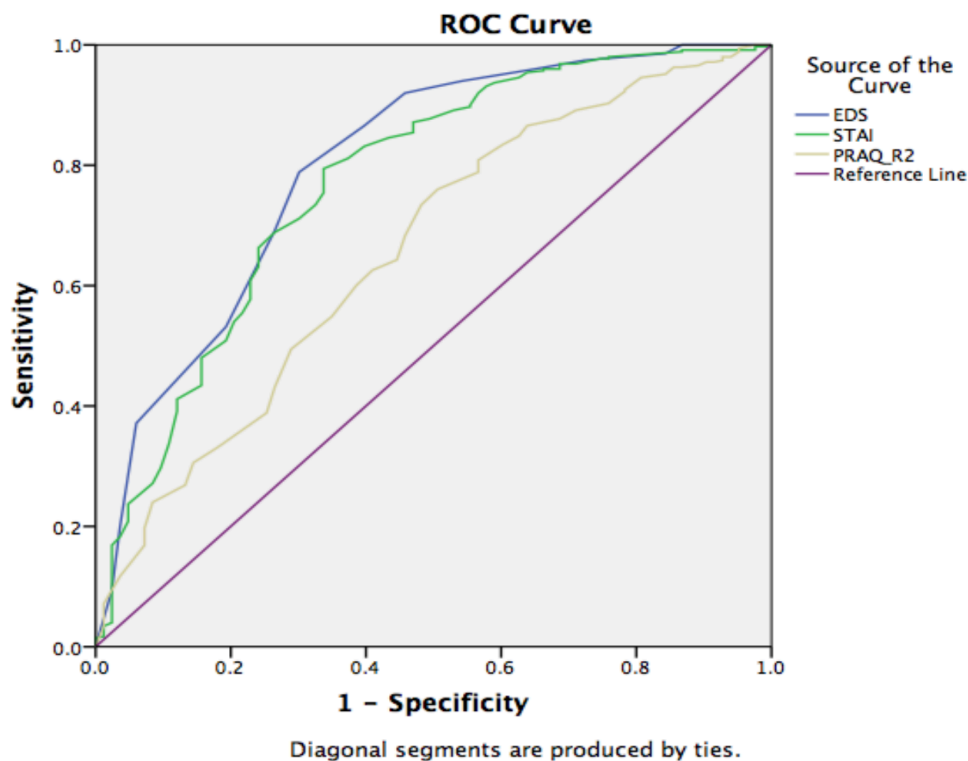

Diagonal segments are produced by ties.
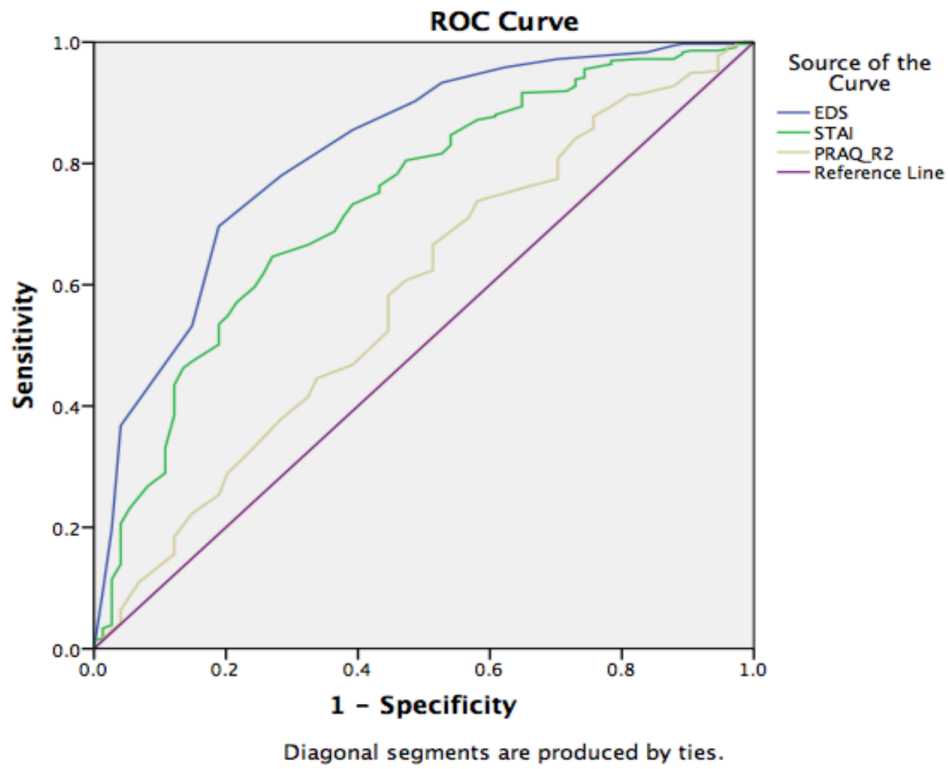
Figure 2. ROC analysis EDS, STAI and PRAQ-R2 at T1, Whooley 2

Figure 3. ROC analysis EDS, STAI and PRAQ-R2 at T2, Whooley 1

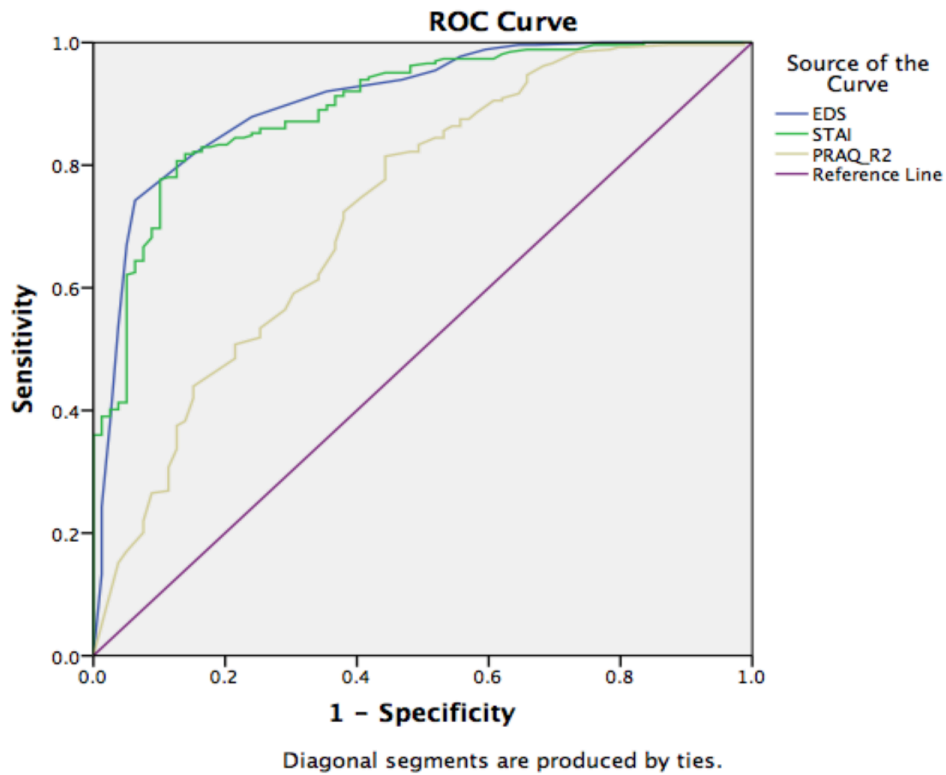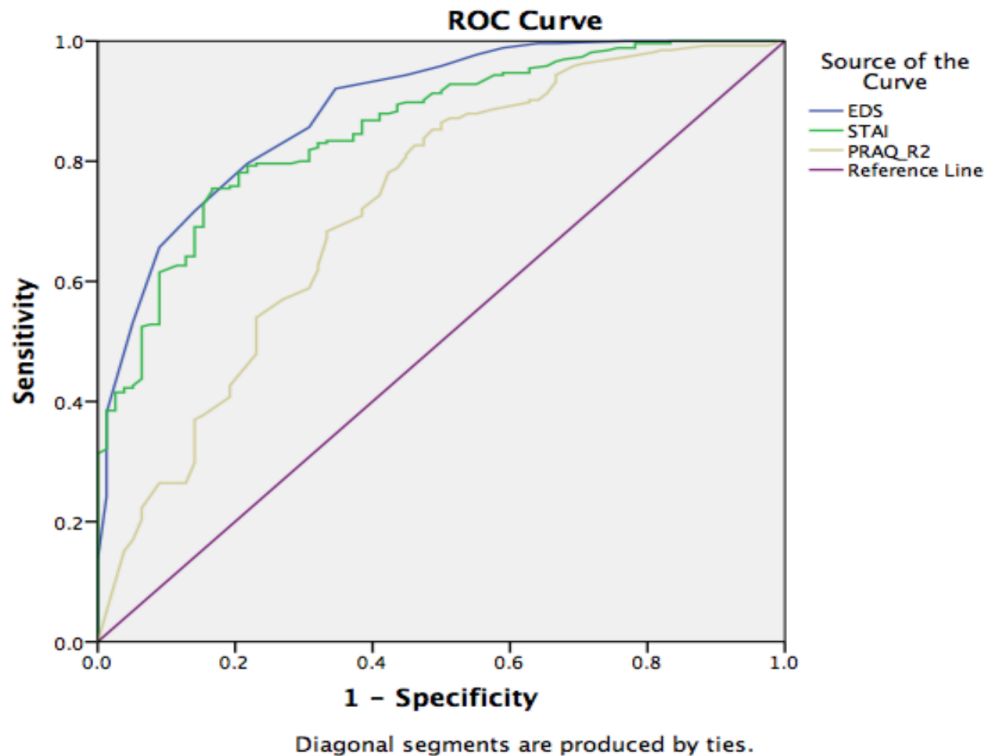Figure 4. ROC analysis EDS, STAI and PRAQ-R2 at T2, Whooley 2

ROC Curve

Source of the Curve
— EDS
— STAI
— PRAQ_R2
— Reference Line

Diagonal segments are produced by ties.

Discussion

This is the first study to offer validation of Whooley questions in Dutch midwifery practice. The Whooley items in this study proved to be more able to report how effective the case-finding questions are in identifying women *without* depression, trait-anxiety and pregnancy-related anxiety (ruling out) rather than how effective these are in identifying women *with* depression, trait-anxiety and pregnancy-related anxiety (ruling in); albeit that the Whooley the items were not very accurate to identify pregnancy-related anxiety. The ability to rule out these psychological constructs would help to substantially reduce the number of women needing more extensive evaluation of their antenatal mental wellbeing. Our study showed that the Whooley items did not only identify depression but also identified trait-anxiety. The value of identifying that something 'is wrong' - regardless whether this is depression or anxiety respectively - facilitating referral for specialist diagnosis, might be of more clinical relevance than the importance of identifying a specific condition at that point which is not within the midwife's remit (DiPietro *et al*., 2004; Fontein-Kuipers *et al*., 2015b). Although we identified high scores, we did not differentiate in levels of severity above the cut-off points. Using different values above the set cut-off point might be of interest for future study as severity can affect the sensitivity and the ability to adequately screen for maternal distress (Lalkhen and McCluskey, 2008; van Stralen *et al*., 2009).

The Whooley items did not accurately identify pregnancy-related anxiety. To our knowledge there are no validation studies with the PRAQ-R2 as the gold standard against which the Whooley items have been compared. Studies have highlighted that pregnancy-related anxiety can be distinguished from general anxiety and may warrant specific clinical attention (Blackmore *et al*., 2016; Huizink *et al*., 2004). It can, therefore, be suggested to add a specific question to assess or screen for pregnancy and birth-related

anxiety, for example: "Are you anxious about the course of the upcoming delivery" (Laursen *et al.*, 2008). Addressing feelings and emotions concerning pregnancy or the forthcoming labour and birth reflects the midwife's unique supportive role. The addition of a pregnancy and birth-related anxiety question would fit the perspective of approaching emotional health of childbearing women as a multi-dimensional concept, indicated by symptoms of different psychological constructs (Fontein-Kuipers *et al.*, 2015a; Fontein-Kuipers *et al.*, 2015b; Fontein-Kuipers *et al.*, 2016).

The interpretation of the results of the Whooley items does not only depend on the accuracy, the sensitivity and specificity, but also on having a history of psychological problems (van Stralen *et al.*, 2009). A percentage of 24.5 of our sample self-reported a history of psychological problems. We are unaware if it were these women in particular that responded positively to either or both of the Whooley items. However, an earlier study among a similar population of Dutch pregnant women showed the significant positive relation between a history of psychological problems and maternal distress (Fontein-Kuipers *et al.*, 2015a). Consideration of a history of psychological problems may be useful for practical utility in midwifery practice, i.e. selective preventive strategies based on a history of psychological problems to identify those women that are more vulnerable for maternal distress.

Granting the recommendation to routinely ask the Whooley questions during booking visits, i.e. early pregnancy (NICE, 2014), our study showed that it might be of worth asking these questions at later stages of pregnancy – considering the increase of sensitivity and positive predictive values from first to third trimester of pregnancy. In addition to this finding, Table 2 shows the increase in positive responses to the Whooley items and the Arroll question. This can be explained because at T2 participants might have been more familiar with the character of the questions, as a repeat measure, and/or women might have given maternal distress more thought during the course of pregnancy. Another interpretation is that women are not keen to disclose their feelings or problems or to ask for help early on in pregnancy when they are firstly engaging with maternity services. They might be more likely to divulge in later pregnancy because they have developed a relationship with the midwife and so feel more comfortable to divulge and potentially less stigmatised. Asking the Whooley questions later in pregnancy might benefit women in their disclosure of maternal distress and might help women to also recognise to themselves that they are struggling further on in pregnancy. An earlier study showed that maternal distress increases during pregnancy in untreated women (Fontein-Kuipers *et al.*, 2016). Based on increased sensitivity, and positive predictive values of the Whooley items as well as increased positive responses to the Whooley items and the Arroll question from first to third trimester, it might be worth considering the utility of these questions as a continuous assessment tool of emotional wellbeing throughout pregnancy (Fontein-Kuipers *et al.*, 2015c). The positive responses to the Arroll question in our study suggest to have some merit. Despite the fact that the Arroll question has been removed from the antenatal and postnatal mental health guideline in the United Kingdom (NICE, 2014; Bosanquet *et al.*, 2015), there may be merit in further consideration and investigation of its value in incorporating the Arroll question in the pathway at a later time-point during pregnancy. Continuous assessment of maternal emotional health during the childbearing period or, at least, revisiting the topic, would both support the woman and the midwife in regarding perinatal emotional wellbeing as an important remit of midwifery care (Fontein-Kuipers *et al.*, 2014; Darwin *et al.*, 2016; McGlone *et al.*, 2016). It would also facilitate differentiation of those women with transient from those with enduring maternal distress (Matthey *et al.*, 2013). Hence, to prepare midwives to effectively use case finding instruments such as the Whooley questions to identify pregnant

women at risk requires education and support of midwives (McGlone *et al.*, 2016). In addition, this approach to identification is not possible without a woman-centered approach or continuity of carer (Hatem *et al.*, 2008; Hodnett *et al.*, 2013). The general literature is all focused on early identification, as are the NICE guidance (2014). There is undoubtedly a group of women for whom that early identification is critical but there is a second group for whom the onset is later (Fontein-Kuipers *et al.*, 2016), reinforcing the value of continuing assessment of women's psychological health across the perinatal period.

We used self-completed questionnaires to collect responses for the Whooley items, while in practice these questions are asked in-person. Whooley questions, when routinely asked face-to-face by midwives, show a lower sensitivity than reported in our study (Darwin *et al.*, 2016; Howard *et al.*, 2018). Questionnaires are sometimes viewed as a 'tick box' exercise but it is striking that where women were asked the Whooley questions online i.e. in writing, there was better discriminatory performance of the items, suggesting that being asked questions face-to-face may inhibit disclosure of problems (Darwin *et al.*, 2016). This might be due to the variation in how questions are being asked in clinical practice or affected by expertise, knowledge and attitude of midwives (Fontein-Kuipers *et al.*, 2015a; McGlone *et al.*, 2016; Howard *et al.*, 2018). It might therefore be of interest to consider presenting the Whooley questions in a written or digital form (Fontein-Kuipers *et al.*, 2015c). Primarily because of the context of disclosure - completion at a time and place suitable for the woman, affording more privacy and time (Darwin *et al.*, 2016; McGlone *et al.*, 2016). Hence, only when a clear referral and care pathway is accessible and available for the woman. An example is the 'WazzUp Mama project' that offers a web-based tailored program, focusing on self-identification and self-disclosure of maternal distress, coping strategies and self-management, help-seeking and resources for support and follow-up care (Fontein-Kuipers *et al.*, 2015c).

The results of our study need to be interpreted in the context of the limitations of the study. The sample in our study contained healthy pregnant women, predominantly with a Dutch ethnicity, all in a relationship, with fairly high levels of education and employment. Further studies with case mix are warranted because interpretation of our study results is limited to similar populations of pregnant women - one could obtain different sensitivities and specificities in different groups or more diverse populations (van Stralen et al., 2009). A methodological limitation of our study concerns internal validity lacking comparison with diagnostic interviews or validating qualitative data of women's accounts. This would offer a richer understanding of the utility of the Whooley questions in practice, but also of women's perspectives of acceptability of the Whooley questions. We would recommend to do this for a future study. We selected the top 15% PRAQ-R2 scores to establish heightened levels of pregnancy-related anxiety. This can be regarded as a rather artificial method to identify women with positive scores – implying the possibility for either underreporting or over reporting. This can affect the representation of those women who do clinically suffer from anxiety directly related to pregnancy or birth. These set cut-off points of pregnancy-related anxiety might have affected the sensitivity values in our study. These rates should only be taken as a guide and further studies replicating this method are warranted, as other studies show other cut-off points than those utilised in our study (Matthey *et al.*, 2013).

5. CONCLUSION

There is currently no policy on routine screening for maternal distress in the Netherlands. The Whooley

questions have shown promise as a screening tool, in specific for depression and trait anxiety – it has good utility as a screening tool. Screening for pregnancy-related anxiety requires specific attention. The Whooley questions as a tool to be used throughout the course of pregnancy with the potential for written or online usage is also encouraging. Further attention and research are required, specifically regarding preventive selective utility of the Whooley questions in midwifery practice, continuous use of the case-finding questions throughout the course pregnancy, utility of written or digital forms for offering the questions, sampling with regard to case mix and attention for methodological issues such as use of cut-off points of the PRAQ-R2. The implementation of the Whooley questions into practice would require education and clinical support for midwives and other healthcare practitioners.

Disclosure of interest
There are no conflicts of interest to declare.

Table 3. Whooley items as indicator for depression (EDS)

	First trimester (T1) n = 433		Third trimester (T2) n= 343	
	Whooley item 1	Whooley item 2	Whooley item 1	Whooley item 2
True positive (n)	23	22	42	43
False negative (n)	9	10	16	15
False positive (n)	60	52	37	35
True negative (n)	341	349	248	250
Sensitivity (95% CI)	71.9% (53.3 to 86.3)	68.8% (50 to 83.9)	72.4% (59.1 to 83.3)	74.2% (61 to 84.7)
Specificity (95% CI)	85% (81.2 to 88.4)	87% (83.3 to 90.2)	87% (82.6 to 90.7)	87.7% (83.3 to 91.3)
Likelihood Ratio	4.8 [3.5 to 6.6]	5.3 [3.5 to 7.5]	5.6 [4 to 7.8]	6 [4.3 to 8.5]
Positive predictive value (95% CI)	27.7% (21.8 to 34.5)	29.7% (23.1 to 37.4)	53.2% (44.7 to 61.5)	55.2% (46.5 to 63.5)
Negative predictive value (95% CI)	97.4% (95.6 to 98.5)	97.2% (95.4 to 98.3)	93.9% (91.1 to 95.9)	94.3% (91.5 to 96.3)
AUC for ROC curve	.80 (.74 to .86)	.82 (.77 to .87)	.85 (.80 to .88)	.85 (.81 to .89)

Table 4. Whooley items as indicator for state-anxiety (STAI)

	First trimester (T1) n = 433		Third trimester (T2) n= 343	
	Whooley item 1	Whooley item 2	Whooley item 1	Whooley item 2
True positive (n)	48	48	39	34
False negative (n)	26	35	10	15
False positive (n)	35	26	40	44
True negative (n)	324	324	254	250
Sensitivity (95% CI)	64.9% (52.9 to 75.6)	57.8% (46.5 to 68.6)	79.6% (65.7 to 89.8)	69.4% (54.6 to 81.8)
Specificity (95% CI)	90.2% (86.7 to 93.1)	92.6% (89.3 to 95.1)	86.4% (81.9 to 90.1)	85% (80.4 to 88.9)
Likelihood Ratio	6.7 [4.7 to 9.5]	7.8 [5.2 to 11.8]	5.9 [4 .2 to 8.1]	4.6 [3.3 to 6.5]
Positive predictive value (95% CI)	57.8% (50 to 66.2)	64.9% (55 to 73.6)	49.4% (41.4 to 57.3)	43.6% (35.7 to 51.8)
Negative predictive value (95% CI)	92.6% (90.3 to 94.4)	90.3% (87.8 to 92.3)	96.2% (93.6 to 97.8)	94.3% (91.6 to 96.2)
AUC for ROC curve	.86 (.82 to .89)	.86 (.83 to .89)	.85 (.81 to .89)	.83 (.78 to .87)

Table 5. Whooley items as indicator for pregnancy-related anxiety (PRAQ-R2)

	First trimester (T1) n = 433		Third trimester (T2) n= 343	
	Whooley item 1	Whooley item 2	Whooley item 1	Whooley item 2
True positive (n)	26	18	30	29
False negative (n)	43	51	25	26
False positive (n)	57	56	49	49
True negative (n)	307	308	239	239
Sensitivity (95% CI)	37.7% (26.3 to 50.2)	26.1% (16.25 to 38.1)	54.6% (40.6 to 68)	52.7% (38.8 to 66.4)
Specificity (95% CI)	84.3% (80.2 to 87.9)	84.6% (80.5 to 88.2)	83% (78.1 to 87.1)	83% (78.1 to 87.1)
Likelihood Ratio	2.4 [1.6 to 3.5]	1.7 [1.1 to 2.7]	3.2 [2.3 to 4.6]	3.1 [2.2 to 4.4]
Positive predictive value (95% CI)	31.3% (23.7 to 40.2)	24.3% (16.8 to 33.8)	38% (30.1 to 46.5)	37.2% (29.3 to 45.8)
Negative predictive value (95% CI)	87.7% (85.5 to 89.6)	85.8% (83.9 to 87.5)	90.5% (87.7 to 92.8)	90.2% (87.4 to 92.4)
AUC for ROC curve	.66 (.59 to .73)	.58 (.51 to .66)	.78 (.74 to .83)	.78 (.73 to .82)

REFERENCES

1. Arroll, B., Goodyear-Smith, F., Kerse, N., *et al*. (2005), "Effect of the addition of a "help" question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study", *BMJ*, Vol. 331 No.7521, pp. 884.

2. Beentjes, M., Weersma, R., Koch, W., *et al*. (2012), "NHG-Standaard Zwangerschap en kraamperiode" (tweede herziening) [NHG-Quality Standard Pregnancy and postnatal period (second revision)]. *Huisarts & Wetenschap*, Vol. 55, pp. 1112–25.

3. Bergink, V., Kooistra, M., van den Berg, *et al*. (2011), "Validation of the Edinburgh depression scale during gestation", *J Psychosom*, Vol. 70, No.4, pp. 385–9.

4. Blackmore, E. R., Gustafsson, H., Gilchrist, M., Wyman, C., O'Connor, T. (2016), "Pregnancy-Related Anxiety: Evidence of Distinct Clinical Significance from a Prospective Longitudinal Study", *Journal of Affective Disorders*, Vol. 197, Jun, pp. 251–8.

5. Boer de, J., Zeeman, K. (2008), *Prenatale verloskundige begeleiding. Aanbevelingen voor ondersteuning, interactie en voorlichting [Antenatal Care. Recommendations for support, interaction and health promotion]*, KNOV, Utrecht, Netherlands.

6. Bosanquet, K., Bailey, D., Gilbody, S., *et al*. (2015), "Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis", *BMJ Open*, Vol. 5 No.12, e008913

7. Darwin, Z., McGowan, L., Edozien, L. (2016), "Identification of women at risk of depression in pregnancy: using women's accounts to understand the poor specificity of the Whooley and Arroll case finding questions in clinical practice.", *Arch Womens Ment Health*, Vol. 19, No.1, pp. 41–9.

8. DiPietro, J., Ghera, M., Costigan, K., Hawkins, M. (2004), "Measuring the ups and downs of pregnancy stress", *Journal of Psychosomatic Obstetrics & Gynecology*, Vol. 25, No.3-4, pp. 189-201.

9. Flahault, A., Cadilhac, M., Thomas, G. (2005), "Sample size calculation should be performed for design accuracy in diagnostic test studies", *J Clin Epidemiol,* Vol. 58 No. 859, pp. 859-62.

10. Fontein-Kuipers, Y., Budé, L., Ausems, M., de Vries, R., Nieuwenhuijze, M. (2014), "Dutch midwives' behavioural intentions of antenatal management of maternal distress and factors influencing these intentions: An exploratory survey", *Midwifery*, Vol. 30 No. 2, pp. 234-41.

11. Fontein-Kuipers, Y., Ausems, M., Budé, L., van Limbeek, E., de Vries, R., Nieuwenhuijze, M. (2015a), "Factors influencing maternal distress among Dutch women with a healthy pregnancy", *Women and Birth*, Vol. 28, No.3, pp. e36-e43.

12. Fontein-Kuipers, Y. (2015b), "Reducing maternal anxiety and stress in pregnancy: what is the best approach?", *Current Opinion in Obstetrics and Gynecology*, Vol. 27 No. 2, pp. 128-32.

13. Fontein-Kuipers ,Y., van Limbeek ,E., Ausems, M., de Vries, R., Nieuwenhuijze, M. (2015c), "Using Intervention Mapping for Systematic Development of a Midwife- Delivered Intervention for Prevention and Reduction of Maternal Distress during Pregnancy", *International Journal of Women's Health and Wellness*, Vol. 1 No. 2.

14. Fontein-Kuipers, Y., Ausems, M., de Vries, R., Nieuwenhuijze, M. (2016), "The effect of Wazzup Mama?! An antenatal intervention to prevent or reduce maternal distress in pregnancy", *Archives of Women's Mental Health*, Vol. 19 No.5, pp. 779-88.

15. Glover, V. (2013), "Maternal depression, anxiety and stress during pregnancy and child outcome: what needs to be done", *Best Practice Research in Clinical Obstetrics and Gynaecology*, Vol. 28 No. 1, pp. 25-35.
16. Goodyer, I., Cooper, P. (2011), "Maternal postnatal depression and the development of depression in offspring up to 16 years of age", *Journal of American Academy Child Adolescent Psychiatry*, Vol. 50 No. 5, pp. 460-70.
17. Hatem, M., Sandall, J., Devane, D., Soltani, H., Gates, S. (2008), "Midwife-led versus other models of care for childbearing women", *Cochrane Database Syst Rev*, 4:CD004667.
18. Heron, J., O'Connor, T., Evans, J,. Golding, J., Glover, V. (2004), "The course of anxiety and depression through pregnancy and the postpartum in a community sample", *Journal of Affective Disorders*, Vol. 80, No.1, pp. 65-73.
19. Hodnett, E. D., Gates, S., Hofmeyr ,G. J,. Sakala, C. (2013), "Continuous support for women during childbirth", *Cochrane Database Syst Rev*, 7:CD003766.
20. Howard, L., Ryan, E., Trevillion ,K., *et al*. (2018), " Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy", *The British Journal of Psychiatry*, Vol. 212, No.1, pp. 50-6.
21. Huizink, A. C., Mulder, E. J., Robles de Medina, P. G,. Visser, G.H., Buitelaar, J. K. (2004), "Is pregnancy anxiety a distinctive syndrome?" *Early Hum Dev*, Vol. 79, pp. 81–91.
22. Huizink, A., Delforterie, M., Scheinin, N., Tolvanen, M., Karlsson, L., Karlsson, H. (2016), "Adaption of pregnancy anxiety questionnaire – revised for all pregnant women regardless of parity: PRAQ-R2", *Archives of Women's Mental Health*, Vol. 19, No.1, pp. 125-32.
23. Ingstrup, K., Schou Andersen, C., Ajslev, A., Pedersen, P., Sørensen, T., Nohr, E. (2012), "Maternal distress during pregnancy and offspring childhood overweight", *Journal of Obesity*, Vol. 05.2012, No. 2, pp. 462845.
24. KNOV (2010), *Advies ontwikkeling wetenschapsdomein fysiologische verloskunde [Advice for Science Development of Physiological Midwifery]*, Utrecht, KNOV, Netherlands
25. Lalkhen, A., McCluskey, A. (2008), "Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia", *Critical Care & Pain*, Vol. 8 No. 6, pp. 221-3.
26. Laursen, M., Hedegaard, M., Johansen, C. (2008), "Fear of childbirth: predictors and temporal changes among nulliparous women in the Danish National Birth Cohort", *BJOG*, Vol. 115, pp. 354–60.
27. Leigh, B., Milgrom, J. (2008), "Risk factors for antenatal depression, postnatal depression and parenting stress", *BMC Psychiatry*, Vol. 16 No.8, pp. 24.
28. Loomans, E., van Dijk, A., Vrijkotte, T., et al. (2013), "Psychosocial stress during pregnancy is related to adverse outcomes: results from a large multi-ethnic community-based birth cohort", *Eur J Public Health*, Vol. 23 No. 3, pp. 485-91.
29. Mann, R., Adamson, J., Gilbody, S. (2012), "Diagnostic accuracy of case-finding questions to identify perinatal depression", *CMAJ*, Vol. 184, No.8, pp. E424–30.
30. Marsay, C., Manderson, L., Subramaney, U. (2017), "Validation of the Whooley questions for antenatal depression and anxiety among low-income women in urban South Africa", *South African Journal of Psychiatry*, Vol. 23 No. a1013.

31. Matthey, S., Valenti, B., Souter, K., Ross-Hamid, C. (2013), "Comparison of four self-report measures and a generic mood question to screen for anxiety during pregnancy in English-speaking women", *Journal of Affective Disorders*, Vol. 148, No.2-3, pp. 347-51.

32. McGlone, C., Hollins Martin, C., Furber, C. (2016), "Midwives' experiences of asking the Whooley questions to assess current mental health: a qualitative interpretive study", *Journal of Reproductive and Infant Psychology*, Vol. 34 No. 4, pp. 383-93.

33. Mennes, M., Stiers ,P., Lagae, L., Van den Bergh, B. (2006), "Long-term cognitive sequelae of antenatal maternal anxiety: involvement of the orbitofrontal cortex", *Neurosci Biobehav Rev*, Vol. 30, No.8, pp. 1078-86.

34. Mulder, E., Robles de Medina, P., Huizink, A., Van den Bergh, B., Buitelaar, J., Visser, G. (2002), "Prenatal maternal stress: effects on pregnancy and the (unborn) child", *Early Hum Dev*, Vol. 70 No. 1-2, pp. 3–14.

35. Murray D, Cox J. Screening for depression during pregnancy with the Edinburgh depression scale (EPDS). J *Reproductive Infant Psychology* 1990; 8:99–107.

36. Nast, I., Bolten, M., Meinlschmidt, G., Helhammer, D. (2013), "How to measure prenatal stress? A systematic review of psychometric instruments to assess psychosocial stress during pregnancy", *Paediatr Perinat Epidemiol*, Vol. 27, No.4, pp. 313–322.

37. NICE (2014), *Antenatal and Postnatal Mental Health: Clinical Management and Service Guidance: Updated edition (update 2017). NICE Clinical Guidelines, No. 192*, National Institute for Health and Care Excellence, London, UK.

38. O'Connor, T.G., Heron, J., Golding, J., Beveridge M, Glover V. (2002), "Maternal antenatal anxiety and children's behavioural/ emotional problems at 4 years", *British Journal of Psychiatry*, Vol.180, No.6, pp. 502-508.

39. Perined (2016), *Perinatale zorg in Nederland 2015 [Perinatal care in the Netherlands 2015]*, Perined, Utrecht, Netherlands.

40. Pop, V., Komproe, I., van Son, M. (1992), "Characteristics of the Edinburgh postnatal depression scale in the Netherlands", *Journal of Affective Disorders*, Vol. 26, No.2, pp. 105–110.

41. Robinson, M., Oddy, W., Li, J., *et al*. (2008), "Pre- and postnatal influences on preschool mental health: a large-scale cohort study", *The Journal of Child Psychology and Psychiatry*, Vol. 49 No. 10, pp. 1118-28.

42. Ruiter, M., Sinnema, H., van de Lindt, S. (2010) *Signaleren en screenen, depressie preventive door beroepsgroepen. Eindverslag versie 1*.0 [Detection and screening, prevention of depression by healthcare professionals. Final report version 1.0], Trimbos-instituut, Utrecht, Netherlands.

43. Son van, M., Verkerk, G., van der Hart, O., Komproe, I., Pop. V. (2005), "Prenatal depression, mode of delivery and perinatal dissociation as predictors of postpartum post-traumatic stress: an empirical study", *Clin Psychol Psychotherapy*, Vol. 12, No.4, pp. 297-312.

44. Spielberger, C., Gorsuch, R., Lushene, R. (1970), *Manual for the state trait anxiety inventory*, Consulting Psychologists Press, Palo Alto, USA.

45. Spijker, J., Bockting, C., Meeuwissen, J., *et al*. (2013), *Multidisciplinaire Richtlijn Depressie [Multidisciplinary Guideline Depression]*, Trimbos-institute Utrecht, Netherlands.

46. Spitzer, R., Williams, J., Kroenke K, *et al*. (1994), "Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study", *JAMA*, Vol. 272 No. 22, pp. 1749-56

47. Stralen van, K., Stel, V., Reitsma, J., Dekker, F., Zoccali, C., Jager, K. (2009), "Diagnostic methods I: sensitivity, specificity, and other measures of accuracy", *Kidney International*, Vol. 75 No. 12, pp. 2157-63. Van der Ploeg, H., Defares, P., Spielberger, C. (1980), *Manual of self-reported questionnaire. A Dutch translation of the Spielberger State-Trait Anxiety Inventory*, Swets and Zeitlinger BV, Lisse, Netherlands.

48. Whooley, M., Avins, A., Miranda, J., *et al.* (1997), "Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med*, Vol. 12, No.7, pp. 439–45.

49. Witteveen, A., de Cock, P., Huizink, A., *et al.* (2016), "Pregnancy related anxiety and general anxious or depressed mood and the choice for birth setting: a secondary data-analysis of the DELIVER study", *BMC Pregnancy Childbirth*, Vol. 16 No.1, pp. 363.