Anxiety measures validated in perinatal populations: A systematic review

Rose Meades, MSc & Susan Ayers, PhD

Rose Meades is a doctoral student in the School of Psychology at the University of Sussex, U.K.

Susan Ayers is a Professor of Maternal and Child Health at City University London, and chartered health psychologist.

Correspondence and request for reprints to:

Rose Meades

School of Psychology, University of Sussex, Brighton, Sussex, BN1 9QH, UK

Tel: 0044 (0)1273 872547

Fax: 0044 (0)1273 678058

Email: R.Meades@sussex.ac.uk

Running title: Anxiety measures validated in perinatal populations

Anxiety measures validated in perinatal populations: A systematic review

Abstract

Background

Research and screening of anxiety in the perinatal period is hampered by a lack of psychometric data on self-report anxiety measures used in perinatal populations. This paper aimed to review self-report measures that have been validated with perinatal women.

Methods

A systematic search was carried out of four electronic databases. Additional papers were obtained through searching identified articles. Thirty studies were identified that reported validation of an anxiety measure with perinatal women.

Results

Most commonly validated self-report measures were the General Health Questionnaire (GHQ), State Trait Anxiety Inventory (STAI), and Hospital Anxiety and Depression Scales (HADS). Of the 30 studies included, 11 used a clinical interview to provide criterion validity. Remaining studies reported one or more other forms of validity (factorial, discriminant, concurrent and predictive) or reliability. The STAI shows criterion, discriminant and predictive validity and may be most useful for research purposes as a specific measure of anxiety. The Kessler 10 (K-10) may be the best short screening measure due to its ability to differentiate anxiety disorders. The Depression Anxiety Stress Scales 21 (DASS-21) measures multiple types of distress, shows appropriate content, and remains to be validated against clinical interview in perinatal populations.

Limitations

Nineteen studies did not report sensitivity or specificity data. The early stages of research into perinatal anxiety, the multitude of measures in use, and methodological differences restrict comparison of measures across studies.

Conclusion

There is a need for further validation of self-report measures of anxiety in the perinatal period to enable accurate screening and detection of anxiety symptoms and disorders.

Key words: anxiety, pregnancy, postpartum period, psychometrics, validation, reliability.

Introduction

Anxiety symptoms and disorders are an important area of research in perinatal populations (Matthey et al., 2003; Ross et al., 2003). Symptoms of anxiety are common in the perinatal period. In a UK study of 8,323 women, 13% of women reported anxiety symptoms at either eight weeks or eight months postpartum and of these two-thirds had experienced anxiety in pregnancy (Heron et al., 2004). A prospective study of 1,507 women expecting their first child found 8.5% of women reported intense anxiety or panic attacks occasionally or often between 6 and 9 months postpartum in a computer assisted telephone interview (Woolhouse et al., 2009). A study using diagnostic interviews with women eight weeks postpartum found 8.2% of women had generalized anxiety disorder (GAD), 2.7% had obsessive compulsive disorder (OCD), 1.4% had panic disorder and 4.1% social anxiety disorder (Wenzel et al., 2005). A further 19.7% of women had sub-syndromal GAD, 5.4% sub-syndromal OCD and 15% sub-syndromal social anxiety. GAD was more prevalent in postpartum women than in the general population, OCD and panic disorder were as prevalent as in the general population, whilst social anxiety was less common in postpartum women (Wenzel et al., 2005).

Although prevalence data of anxiety disorders is limited, a review tentatively suggests that rates of OCD and GAD are higher in perinatal populations than in the general population, whilst rates of panic disorder and posttraumatic stress disorder (PTSD) are comparable (Ross and McLean, 2006). There is also preliminary evidence that a further proportion of women experience clinically significant anxiety symptoms that do not fit into DSM-IV anxiety diagnoses (Phillips et al., 2009). In a sample of women with unsettled infants, equal numbers of women were diagnosed by clinical interview with an anxiety disorder not otherwise specified (ADNOS; 10.8%) as were diagnosed with GAD (10.8%) (Phillips et al., 2009). None of the women diagnosed with ADNOS had primary symptoms related to OCD, social anxiety, specific phobias or panic disorder indicating that anxiety disorders in postnatal women may not be optimally defined. Furthermore, the occurrence of pregnancy-specific anxiety that needs to be treated as a relatively distinct syndrome has been proposed (Huizink et al., 2004). State anxiety explains a small part of important fears of giving birth, having a disabled child, and concerns about a woman's appearance when she is pregnant (Huizink et al., 2004). Pregnancy specific stress has

predicted poor birth outcomes better than state anxiety and perceived stress (Lobel et al., 2008). Thus in both antenatal and postnatal women, general measures of anxiety may not be appropriate screening tools.

Anxiety disorders and subthreshold anxiety symptoms can be detrimental to the mother-infant relationship (Zelkowitz and Papageorgiou, 2005). Prenatal anxiety has been linked with intrauterine artery resistance leading to low birthweight (Teixeira et al., 2009; Field et al., 2010), low apgar scores (Berle et al., 2005), suboptimal changes in fetal heart rate and motor activity (DiPietro, 2010) and complications of labour and delivery (Johnson and Slade, 2003). Longer term effects of high antenatal anxiety can include externalizing problems and self-reported anxiety at ages 8-9 (Van den Bergh and Marcoen, 2004) and high impulsivity and lower scores on cognitive tests at ages 14-15 (Van den Bergh et al., 2005). History of an anxiety disorder before pregnancy is a greater risk factor for postnatal anxiety or depression than a history of depressive disorder (Matthey et al., 2003). Additionally, anxiety disorders *during* pregnancy predict postnatal depression symptoms independent of antenatal depression. One study shows that women with anxiety disorders in pregnancy are nearly three times more likely to present with postnatal depression (Sutter-Dallay et al., 2004). A large prospective study (n = 12,361) confirms that antenatal anxiety is one of the strongest predictors of postnatal depression indicating that antenatal screening for anxiety may be helpful in reaching women at risk of postnatal depression (Milgrom et al., 2008). A recent review indicates that postnatal maternal anxiety is associated with deleterious outcomes for children such as alterations in emotional and conduct problems, and adverse effects on cognitive and social developments, although effects on infant temperament are inconclusive (Glasheen et al., 2010).

As in other clinical populations, there is substantial comorbidity between anxiety and depression in perinatal women. Studies using clinical interviews to diagnose disorders show that up to 50% of women with affective disorders have comorbid anxiety and depression. For example, a study of anxiety and depression at 36 weeks gestation, 6 weeks postpartum and 16 weeks postpartum found that nearly 50% of clinically depressed women experienced clinically significant comorbid anxiety (Ross et al., 2003). Depending on the domain of anxiety being considered, Wenzel et al. (2005) found

that 10-50% of women reporting anxiety symptoms also endorsed depressive symptoms. In two samples of women interviewed at 6 weeks postpartum Matthey et al. (2003) found 16.2% and 10.4% of women had anxiety disorders only, whereas 5.6% and 2.6% had major or minor depression only, and 4.2% and 2.1% had comorbid depression and anxiety highlighting the importance of screening for both depression and anxiety.. These studies all used clinical interviews to diagnose anxiety and depression.

Given the high prevalence (but also variation) of perinatal anxiety, it is important to identify valid questionnaire measures of anxiety that can be used in research and clinical practice. Two broad approaches can be taken: (i) to use measures of anxiety developed with other populations and validate them for use with perinatal women; or (ii) to develop specific measures of perinatal anxiety. Research and screening for perinatal depression has taken the latter approach, as illustrated by the Edinburgh Postnatal Depression Scale (EPDS; Cox et al., 1987) which is the most widely used screening tool used for postnatal depression (Boyd et al., 2005).

The creation of the EPDS and other measures of depression specific to the perinatal period derived partly from the inappropriateness of general depression measures to perinatal women, largely because of the inclusion of items measuring somatic symptoms. Specific depression measures such as the EPDS were therefore developed and validated for use in the perinatal period. However, the utility of the EPDS appears to vary in different samples (Gibson et al., 2009) and there is evidence that the EPDS has three items that also measure anxiety. Some propose it therefore contains an unofficial anxiety subscale, but this only contains three items which is likely to be insufficient to capture the complexity of anxiety as a construct. Muzik et al. (2000) found that women with anxiety disorders had significantly lower scores on the EPDS than women with major depressive disorder. These women would therefore not be identified using only the EPDS. Thus, if the EPDS is the only measure used to screen perinatal women, it will be unclear whether the anxiety symptoms are features of depression or are a separate clinical entity (Ross et al., 2003). There is also evidence that the depression subscale of the EPDS correlates more highly with other measures of anxiety than the anxiety subscale does, indicating poor convergent validity (Brouwers et al., 2001).

With the exception of measures of pregnancy-specific worries or anxiety, no specific measures of anxiety have been developed for use in perinatal populations. Most measures used were originally designed for use in generic samples. Validity of any self-report measure depends on recalibration for the population under study (Geisinger, 1994). Self-report measures constructed for use with one population may produce flawed results in another population because the distribution of the variable will differ according to a number of factors including culture, development and time of measurement in the perinatal period. Normative values and cut-offs may also vary in samples that differ from the original sample. Therefore an evaluation must be made as to whether the sample with whom the test was designed is sufficiently similar to the test sample to ensure little variability in functioning of the self-report measure (Myers and Winters, 2002). Self-report measures created for use in populations in western countries are often validated for use in different languages and cultures (Prince, 2008). However measures designed for use in general populations are not as frequently validated for use in perinatal populations despite the unique nature of this period.

Research into perinatal anxiety uses a variety of self-report measures including the State Trait Anxiety Inventory (STAI; Spielberger et al., 1970), Crown-Crisp Experiential Index (Crisp et al., 1978) and Beck Anxiety Inventory (Beck et al., 1988). These general measures are rarely validated for use in perinatal populations which could lead to erroneous interpretation and incomparable data. A variety of different cut-off points on measures have also been used to indicate severe anxiety or disorder. For example the STAI has been used in perinatal samples with cut-offs of >40 (Grant et al., 2008) to >45 (Moss et al., 2009) or >48 (Field et al., 2010) to indicate disorder, without prior validation in this population.

In sum, there is as yet no measure specifically devised to screen for anxiety disorders or symptoms in perinatal populations. Psychometrically robust measures of anxiety are needed for screening and research purposes. This paper therefore aims to systematically identify and review measures of anxiety that have been validated in perinatal populations.

Methods

Data Search

A systematic search of the following computerised databases was undertaken: SCOPUS V.4 (Elsevier), MEDLINE (CSA), PsycINFO (CSA) and the Cumulative Index to Nursing and Allied Health Literature (CINAHL) for the time period from inception of the database to 30 September 2010. The following search terms were used to search all databases: *anx**, *pregnan** or **natal* or **partum*, *diagnos** or *screen* or *questionnaire* or *tool* or *scale*. In addition the reference lists of included papers were searched.

Inclusion / exclusion criteria

Inclusion criteria were that published studies i) were written in English; ii) examined the reliability and/or validity of a self-report anxiety measure; iii) the sample were antenatal or postnatal women (up to one year postpartum). A number of studies validated more than one measure on the same sample resulting in more than one publication. These papers are included with a note indicating the multiple use of the sample. To enable comparison, all measures were originally devised in English.

Data quality

Quality of studies was assessed using a combined checklist based on the quality assessment of diagnostic accuracy studies (QUADAS) (Whiting et al., 2003) and a checklist developed by Mirza and Jenkins (2004). Criteria were assessed as present or absent on 11 dimensions (range 0 – 11): 1) explicit study aims; 2) adequate sample size; 3) sample described in sufficient detail; 4) sample representative of population receiving test in practice; 5) clear inclusion and exclusion criteria; 6) use of appropriate reference standard; 7) reliability of measure reported 8) validity of measure investigated; 9) specification of dropouts and withdrawal of participants; 10) adequate description of data and, 11) discussion of generalisability. The studies were then given a total score of quality with the highest possible being eleven. Most studies were of good quality with 22 (73%) having a score of 8 or more. Inter-rater reliability was checked for four (13%) studies and agreement across all 11 dimensions for those studies was very high (mean agreement across studies was 97%).

Validity

Various types of validity data were extracted. *Criterion validity* is inferred by comparing the sample's score on the index measure with a more objective measure of the same construct. Criterion validity should be established against an existing gold standard (Greenhalgh, 1997). For determining anxiety disorders, the gold standard is defined as a diagnostic clinical interview (Gibson et al., 2009). Validity is inferred by a measure correctly identifying most people with the disorder (high sensitivity) and correctly excluding most people without the disorder (high specificity). In line with a review of measures to detect postnatal depression, the following criteria were applied to this type of psychometric data: >.90 as excellent, .90-.80 as good, .70-.50 as moderate, .50-.30 as low, and <.30 as poor (Boyd et al., 2005). Overall misclassification rate is the proportion of people incorrectly diagnosed.

Available data were also extracted for *concurrent validity*, the ability of a test to correlate positively with measures of the same construct or negatively with measures of an opposing construct. Cohen's (1992) criteria were used to evaluate effect size: $r = 0.1$, small; $r = 0.3$, medium; $r = 0.5$, large. *Discriminant validity* (the ability of the measure to discriminate between groups differing in levels of anxiety) and *factorial validity* (the ability of the construct to provide a clear factor structure) was also extracted.

Reliability

The reliability of measures was evaluated by examining *internal consistency* and *test-retest reliability*. *Internal consistency* represents the extent to which different items of the same construct are scored similarly. For a scale to be reliable a minimum Cronbach's alpha coefficient of 0.70 is recommended (Kline, 2000). Reliability is a measure of the properties of a test within a particular sample rather than a property of a test per se and therefore should be computed in each sample (Vacha-Haase et al., 2000). *Test-retest reliability* (the ability of a test to yield consistent scores over time) should be greater than .80 over one to two weeks or .70 or greater over one month (Boyd et al., 2005). As it is expected that anxiety levels fluctuate throughout pregnancy and the postpartum, test-retest reliability is only relevant to trait anxiety which should remain stable over time (Spielberger et al., 1970).

Results

Methods of validation and assessment of reliability

Thirty studies were included. Studies often reported more than one type of validity. One study (Aboidun, 1994) validated two measures. Eleven studies (37%) used clinical interviews for criterion validity; 10 (33%) reported concurrent validity; 9 (30%) reported discriminant validity and 8 (27%) reported factorial validity. Three studies (10%) reported the validity of different scoring methods of a measure. Eleven studies (37%) reported internal consistency and two studies (7%) reported test-retest reliability. Eighteen studies (60%) did not report any reliability data. The heterogeneity of the studies and types of validity reported precluded any scope for meta-analysis in this review. An overview of included studies is given in Table 1.

- Insert Table 1 about here -

Self-report measures

Thirteen studies validated the General Health Questionnaire (GHQ), five studies validated the State-Trait Anxiety Inventory (STAI) and three studies validated the Hospital Anxiety and Depression Scales (HADS). A further nine studies reported a different measure each. Following is a description of each measure, a summary of reported validity and reliability, and an indication of its utility in perinatal samples.

Hospital Anxiety and Depression Scales – Anxiety subscale (HADS-A)

The HADS (Zigmond and Snaith, 1983) consists of an anxiety subscale (HADS-A) and a depression subscale (HADS-D), each comprising seven items to assess presence or absence of symptoms over the last week, on a four-point scale (0-3). Scores of 0-7 are considered 'normal,' 8-10 is considered suggestive of disorder, and 11+ indicates probable presence of a mood disorder (Snaith, 2003). A cut-off of 8 has optimal sensitivity and specificity in non-perinatal samples (Bjelland et al., 2002). The scale was designed to include minimal reference to somatic symptoms (Snaith, 2003). A number of items on the HADS-A scale may have a different relevance to pregnant or postpartum women

compared with other populations. For example 'I can sit at ease and feel relaxed' may be less likely in pregnant and postpartum populations.

Three studies assessed the validity of the HADS-A throughout pregnancy using the recommended cut-off of 8 to indicate possible anxiety disorders. No studies assessed the HADS-A in the postpartum. Only one study (Abiodun, 1994) reported validity coefficients. Sensitivity and specificity were particularly high (92.9% and 90.0%) and the misclassification rate was low (9.6%). This cut-off led to a high prevalence of possible cases of anxiety in the UK (36-56%) and Uzbekistani samples (38-42%) (Jomeen and Martin, 2004; Karimova and Martin, 2003) but low prevalence in Nigeria (5.8%; Abiodun, 1994).

One study reported concurrent validity. The HADS-A correlated more highly with the EPDS (r = 0.73) than with the HADS-D (r = 0.46) indicating a strong association with postnatal depression (Jomeen and Martin, 2004). Test-retest correlation between 12 and 34 weeks gestation and was low (r = 0.31) as may be expected over the course of pregnancy (Karimova and Martin, 2003). Internal consistency across the studies ranged between 0.62 and 0.78 (Karimova and Martin, 2003; Jomeen and Martin, 2004).

The factor structure of the HADS is unclear. In non-perinatal samples there is conflicting evidence of two, three or four factors (Bjelland et al., 2002). Two studies with perinatal women examined the factor structure of the HADS. Jomeen and Martin (2004) carried out principal components analysis (PCA) and confirmatory factor analysis (CFA). Both methods identified three factors: one anxiety (5 items) and two depression factors, although two anxiety items did not load on the anxiety factor. These factors only accounted for around half of the variance in the scale, as indicated by the PCA (r^2=.52) and fit indices of the CFA. Another study comparing the HADS structure for UK and Uzbekistani women at 12 and 34 weeks gestation found between two and five factors optimally fitted the data dependent on the time of gestation and nationality of group (Karimova and Martin, 2003).

State Trait Anxiety Inventory (STAI)

The STAI (Spielberger et al., 1970) consists of two subscales each with 20 items. The state subscale measures anxiety *at this moment* (related to a specific situation or time period). The trait subscale measures relatively stable individual propensity to respond with elevated anxiety i.e. *as you generally feel*. Respondents endorse items on a four-point scale (1-4). Some items such as 'I tire quickly' and 'I feel rested' may be confounded by pregnancy or postpartum factors.

Six studies reported use of the STAI in perinatal populations. Two of these used the same sample (Barnett and Parker, 1985, 1986). One paper (Grant et al., 2008) reported validity coefficients. A cut-off >40 on both state and trait scales yielded optimal sensitivity (80.95%), specificity (79.75%), positive predictive value (51.50%) and negative predictive value (94.00%) to determine cases of anxiety in the third trimester of pregnancy. No studies have conducted a factor analysis on STAI in perinatal populations. Norms available from included studies are given in Table 2.

Both the state and trait anxiety subscales are highly correlated with measures of depression. The subscales correlated more highly with the EPDS than with the BDI at 14 and 30 weeks postpartum. Correlations between the state subscale and a measure of anxiety (BAI) were similar to that observed with depression (Stuart et al., 1998).

Barnett and Parker (1985, 1986) recommend cut-offs of high (≥ 40), moderate (32 - 33) or low (≤ 25) anxiety on the basis of mean trait scores of a sample of 94 primiparae (M =33.1, SD = 8.1; Barnett and Parker, 1985). Discriminant validity of these thresholds was shown by statistically significant differences in scores between groups on a number of self-report measures (Costello-Comrey trait anxiety and depression scales, Beck Depression Inventory, Eysenck Neuroticism Scale). More highly anxious women endorsed significantly more life events in the previous 12 months, and more distress associated with those life events, lower parental confidence and lower confidence in their maternal role (all ps <.05). Barnett and Parker (1986) reported further discriminant validity through inspection of hospital files where nursing staff considered a significantly higher percentage of women in the high anxiety group to have 'mood problems' than in the moderate or low anxiety groups. The interviewer, who was blind to group allocation, rated women with anxiety in the anticipated direction, but

differences between groups were not significant. Interviewer ratings of increased arousal and unassertiveness also significantly differentiated between anxiety groups.

Predictive validity of the STAI has been demonstrated in postnatal samples. A trait anxiety score of >40 was associated with a six-fold increase in postnatal anxiety disorders (odds ratio = 6.44, CI = 1.28 - 32.28) and depression (odds ratio = 6.12, CI = 1.37 - 27.41) whilst antenatal state anxiety did not predict postnatal anxiety or depressive disorders (Grant et al., 2008).

Test-retest reliability would be expected to be low for the state scale of the STAI, which should measure momentary fluctuations in anxiety, and high for the trait anxiety, which should remain stable. The latter is not the case in perinatal samples. Hundley et al (1998) found test-retest correlations of between 0.37 - 0.85 on the trait scale. In addition, women all had lower trait scores after birth than in pregnancy. In Barnett and Parker's (1985) sample, trait scores decreased by 2 - 10% between 3 days and 12 months postpartum. This has also been found in a Greek sample of mothers, 10.3% of whom had substantially lower trait scores at 3 months postpartum than at 2-3 days postpartum (Giakoumaki et al., 2009). In general, internal consistency is good, ranging from 0.91 – 0.95 for the state subscale of the full version (Marteau and Bekker, 1992; Grant et al., 2008) and 0.82 for the 6 item version (Marteau and Bekker, 1992). Reported trait scale internal consistency was 0.96 (Grant et al., 2008).

Six and four item versions of the state scale of the STAI have been constructed and validated for use in pregnant populations (Marteau and Bekker, 1992). Concurrent validity was demonstrated by pro-rating scores from the shortened versions and statistically testing for differences between original and pro-rated scores. The shortest scale score that did not differ significantly to the pro-rated total score consisted of 6 items correlating >.90 with scores on the full scale.

- insert Table 2 about here -

General Health Questionnaire (GHQ)

The GHQ (Goldberg, 1972) is a widely used measure of non-psychotic psychopathology, used to screen for cases requiring further psychiatric consultation. It asks how the respondent has felt recently. Answers are given on a four-point scale with higher scores indicating increased likelihood of disorder. The GHQ has four different versions (60-item, 30-item; 28-item and 12-item versions) and can be scored in four different ways (Likert scoring 0 – 3; modified Likert scoring 0-0-1-2; C-GHQ scoring 0-1-1-1 for negative items (indicating illness) and 0-0-1-1 for positive items (indicating health); or dichotomous 0-0-1-1 scoring to indicate healthy and ill responses). The shorter versions of the GHQ do not result in greatly reduced sensitivity and specificity in non-perinatal populations (Ayers, 2001). Only the GHQ-28 has a specific anxiety subscale. Thirteen studies evaluated versions of the GHQ in perinatal samples. None of these used the 60 item version. Validity coefficients are shown in Table 3.

GHQ-30

Seven studies (comprising four different samples) used the GHQ-30. The GHQ-30 is the most widely validated version in non-perinatal populations and yields an overall total score. The recommended cut-off for identifying cases of psychiatric disorder is 4/5 for GHQ scoring and 23/24 for Likert scoring (GL Assessment, 2010).Cut-offs used in perinatal studies ranged from 5 in a Nigerian sample (Abiodun, 1993) to 7/8 in a Japanese sample (Kitamura et al., 1989, 1994a, 1994b). Specificity ranged from 71 - 89% across studies. Sensitivity was mostly in the range of 77 - 83%. However a Japanese study in pregnancy and at different time points postpartum found no cut-off yielded satisfactory sensitivity (Kitamura et al., 1994a). In addition to validating the GHQ-30, Nott and Cutts (1982) reported validity coefficients for a moderated version with two items that may be confounded by perinatal circumstances removed. The modified version increased specificity and decreased overall misclassification.

Two studies performed discriminant function analyses. Kitamura et al (1989) showed that only 13 out of 30 items significantly contributed to the power of the measure to differentiate cases from non-cases. Items that did not discriminate included: ‘Have you been getting out the house as much as

usual'; 'Have you been managing to keep yourself busy and occupied' and 'Have you spent much time chatting to people'. However Nott and Cutts (1982) found that 28 out of 30 items were related to 'caseness'. Their discriminant function analysis resulted in a combination of 12 items which if used together resulted in high sensitivity (98%) but low specificity (68%) and a high misclassification rate (27%).

The GHQ scoring method provided the greatest discriminatory power between cases on the GHQ-30 (Kitamura et al., 1993).

GHQ-28

Three studies validated the GHQ-28. The GHQ-28 (Goldberg and Hillier, 1979) contains four subscales measuring depression, somatic symptoms, anxiety/insomnia and social dysfunction. The factors do not have individual cut-off scores and caseness is determined from an overall score. Sensitivity and specificity ranged from 75 - 85% and were in keeping with a large study (n=5438) which also showed that cut-off points differed depending on culture and geographical locations (Goldberg et al., 1997). Across all studies sensitivity and specificity for all scoring methods were within 75 - 85% which is comparable with the overall sensitivity (79.7%) and specificity (79.2%) in Goldberg et al. (1997). Positive predictive value ranged from 46 - 53% (Aderibigbe and Gureje, 1992).

Correlations between the somatic symptom scale and clinical interview scores and between the anxiety subscale and clinical scores were significant but low (*r*s = 0.26 to 0.29; Aderibigbe and Gureje, 1992). Social dysfunction and depression subscales did not correlate significantly (Aderibigbe and Gureje, 1992). Discriminant function analysis of the subscales showed that the social dysfunction subscale did not make a significant contribution to the GHQ-28's ability to discriminate cases from non-cases (Aderibigbe and Gureje, 1992).

The factor structure of the GHQ-28 in an antenatal Nigerian population (Aderibigbe et al., 1996) differed to that found in UK primary care populations (Goldberg, 1979) although it is unclear whether this is due to cultural differences, the experience of pregnancy, or an interaction between the two. In

the Nigerian sample, principal components analysis resulted in four factors explaining 40.1% of the variance in the data. Social dysfunction accounted for most variance (19%). Anxiety items were hypothesised to load onto a single factor but were divided between somatic-anxiety (9.7% of variance) and depression-anxiety (6.3% of variance) subscales. Severe depression accounted for 5.1% of the variance (Aderibigbe et al., 1996).

Two studies considered scoring methods of the GHQ in perinatal samples. One study found that the C-GHQ method of scoring resulted in higher sensitivity and specificity than the standard GHQ scoring in Nigerian women in the late second or early third trimester (Aderibigbe and Gureje, 1992). The other study of a UK sample in the late first trimester suggested that the number of cases identified depended on the scoring method used with the standardized cut-off of 4/5 classifying more women as cases than the Likert scoring (Swallow et al., 2003).

GHQ-12

Three studies validated the GHQ-12 in perinatal samples. The GHQ-12 was compiled by removing items from the GHQ-60 that relate to physical symptoms and then selecting equal numbers of health present and health absent items that discriminate well between cases. The GHQ-12 has the advantage of being free of somatic symptoms, although it cannot distinguish between mood states as it gives a total score only and is not divided into subscales. It is quick to administer (approximately two minutes) and sensitivity and specificity for anxiety disorders were good in antenatal and postnatal populations in the two studies examining this (>0.80). The standardized threshold to indicate a case is 1/2 for GHQ scoring and 11/12 for Likert scoring (GL Assessment, 2010). Cut-off scores had to be increased from the suggested standardized thresholds in order to obtain satisfactory validity coefficients. After birth, the GHQ-12 detected 69.6% of pure anxiety cases and 86.9% of depression with or without comorbid anxiety using a cut-off of 4/5 (Navarro et al., 2007). Post-hoc tests showed that the GHQ-12 found significant differences in scores between cases and non-cases, but did not find differences between diagnostic groups of anxiety, depression, comorbid anxiety and depression, or adjustment disorder.

The EPDS and GHQ-12 showed good concurrent validity (r = 0.80) (Navarro et al., 2007). An effect of scoring method on case identification was found with the GHQ-12 (Martin and Jomeen, 2003). GHQ scoring resulted in 34.5% cases, C-GHQ scoring in 52.7% cases and Likert scoring in 45.5% of cases. The scores from the C-GHQ method resulted in significant main and interaction effects in an analysis of variance (ANOVA) but no effects were found with the Likert and GHQ methods in this sample of women with pre-labour rupture of membrane at term (Martin and Jomeen, 2003). The difference in number of cases identified and interpretation of ANOVA results depending on scoring method raises concerns about the reliability of the GHQ-12 in late pregnancy/early postpartum or at least highlights that scoring methods need to be validated against clinical interviews in this population. Internal consistency ranged from 0.81 - 0.95 (Martin and Jomeen, 2003). Validity coefficients are shown in Table 3. Only one paper reported normative data (Navarro et al., 2007).

- insert Table 3 about here -

Other measures

The Personal Disturbance Scale (DSSI/sAD; Bedford and Foulds, 1978) is a 12-item self-report measure of anxiety and depression. One study validated the DSSI/sAD against a clinical interview at four time points in pregnancy and postpartum (Cox and O'Connor, 1983). This study found that the standardized cut-off point of 6 was not appropriate for pregnant women: 13 women out of 230 scored ≥6 and of these only three had clinically significant psychiatric illness but not anxiety or depression. The scale did not identify two women with phobic neuroses. Increases in anxiety and depression at the first postpartum time point which were detected by clinical interview and visual analogue scales were not identified by the DSSI/sAD (Cox and O'Connor, 1983). Factor analysis in a non-perinatal population suggests the scale may be measuring one latent factor rather than two distinct anxiety and depression subscales (Shevlin et al., 1998).

Two studies examined variations of the Symptom Checklist 90 Revised (SCL-90-R; Derogatis, 1992). The SCL-90-R comprises 90 items covering 9 primary dimensions including anxiety, phobic anxiety, somatisation, depression and obsessive compulsive scales, each measured on a five point scale. In a

small sample of 3 or 6 month postpartum women the SCL-90-R anxiety subscale scores distinguished women with anxiety disorders from women with no psychiatric diagnosis but did not distinguish anxiety cases from depression cases in post-hoc tests (Muzik et al., 2000). Women with major depressive disorder also had higher scores on the anxiety subscale than women with anxiety disorders although no women included in the analysis had comorbid anxiety and depression (Muzik et al., 2000).

The Hopkins-Symptom Checklist 25 (Derogatis et al., 1974) is a self-report inventory of 10 anxiety and 15 depression items derived from the SCL-90. A factor analysis in a large sample of pregnant HIV positive Tanzanian women yielded four factors representing depression-anxiety, anxiety, psycho-physiological and physiological constructs rather than two clear factors (Kaaya et al., 2002). This effect could be due to pregnancy or diagnosis of HIV. Optimal cut-off points and validity coefficients were reported for detecting major depressive disorder only. The anxiety subscale showed good internal consistency (Cronbach's alpha = 0.85).

One study suggested that a single item 'Do you feel very anxious?' could be used to screen for anxiety in antenatal clinics where time is constrained (Sagrestano et al., 2001). This item correlated with standardized scores on the STAI showing a small to medium effect size (r = 0.24). Women who answered 'yes' to feeling anxious had significantly higher scores on the STAI than those who had answered no. The same situation was true for single items measuring depression and social support but single items measuring stress, relationship conflict, verbal abuse or violence from partner did not reliably differentiate between women experiencing high and low levels of these constructs. The questions were asked as part of an interview rather than a self-report questionnaire which may have influenced women's decision to disclose such information (Sagrestano et al., 2001).

The Pregnancy Anxiety Scale (PAS; Levin, 1991) is the only measure developed specifically for pregnant women. It measures pregnancy-specific anxiety with 13 yes/no items. Confirmatory factor analysis showed a clear structure of three latent factors measuring anxiety about being pregnant,

anxiety about childbirth and anxiety about hospitalization with good model fit indices. However data about pregnancy anxiety was collected postpartum and no further studies have validated this measure.

The Depression Anxiety Stress Scales (DASS; Lovibond and Lovibond, 1995) has three subscales designed to maximally discriminate between depression, anxiety and stress in the last week (Antony et al., 1998). The DASS excludes somatic items such as sleep disturbance, lack of energy and poor concentration which may not be valid markers for pregnant or postnatal women. The DASS is available in 42 or 21 item versions. The DASS-21 appears to measure anxiety without a large overlap with depression. In a study of 325 postnatal women, 240 women had EPDS scores <9 (a negative screen). Yet the DASS-21 identified eight of these women (3% of the total sample) as anxious and a further 10 women (4% of the total sample) as stressed. Internal consistency was acceptable for all subscales: 0.84, depression; 0.77, anxiety and 0.86, stress (Miller et al., 2006).

The Kessler 10 (K-10; Kessler et al., 2002) is a short screening scale comprising 10 items designed to assess level of general psychological distress in the past four weeks. Scores can range from 10 (no distress) to 50 (severe distress). In a sample of women in early pregnancy the K-10 identified 69 - 76% of diagnosed cases of anxiety disorders (Spies et al., 2009). Sensitivity ranged from 0.50 to 1.00 and specificity from 0.75 - 0.98 dependent on the anxiety disorder in question. The scale performed best at detecting social anxiety disorder (sensitivity 1.00, specificity 0.98) but validity coefficients were lower for panic disorder (sensitivity 0.50, specificity 0.98) and PTSD (sensitivity 0.50, specificity 0.80). Numbers of women with anxiety disorders were low, limiting power to show representative sensitivity and specificity. However, the K-10 has shown robust psychometric properties in non-perinatal samples and had substantially better validity coefficients than the GHQ-12 in discriminating cases of anxiety and mood disorders in the Australian National Survey of Mental Health and Well-being (Kessler et al., 2002; Furukawa et al., 2003). A Dutch study improved the ability of the K10 to identify anxiety disorders by adding five further items pertaining to GAD, social anxiety disorder, panic and agoraphobia (Donker et al., 2010).

Finally, two new measures have attempted to broaden the content of anxiety and psychological distress measures. Firstly, the Inventory of Depression and Anxiety Symptoms (IDAS; Watson et al., 2007) is a new multi-dimensional measure of symptoms of anxiety and depression, containing 64 items which yielded 10 factors in multiple samples relating to symptoms of: suicidality, lassitude (fatigue, lack of energy, hypersomnia), insomnia, appetite loss, appetite gain, ill temper, well-being, panic, social anxiety, and traumatic intrusions as well as two further scales assessing general depression scale and dysphoria. Responses refer to the previous two weeks and are scored on a scale from 0-3. In a large sample (n=832) of women less than 4 months postpartum, the 10 initial symptom scales all loaded onto a higher order factor of general distress with loadings ranging from 0.38 (appetite loss) to 0.76 (social anxiety). Dysphoria had loading of 0.91. These 11 factors accounted for 92% of the variance in the data. All IDAS scales correlated significantly with the EPDS and with clinician ratings on the Hamilton Rating Scale for Depression (Watson et al., 2007). Internal consistency ranged from 0.74 (suicidality) to 0.91 (general depression). The IDAS remains to be validated against a clinical interview in perinatal populations but has shown strong correlations in a psychiatric sample particularly between SCID diagnoses of panic disorder and IDAS Panic, posttraumatic stress disorder and IDAS Traumatic Intrusions, and social anxiety disorder and IDAS Social Anxiety (Watson et al., 2008). However, GAD correlated most strongly with IDAS General Depression and Dysphoria, and OCD diagnoses did not show differentially high correlations with any IDAS subscales.

Secondly, an unnamed measure was constructed through an internet survey of 142 questions derived from case descriptions of postnatal psychiatric disturbance and DSM-IV-R, with participants asked to retrospectively rate the frequency with which they experienced symptoms in the first 30 days after birth (Marrs et al., 2009). This measure was not intended for screening purposes but provides evidence from factor analysis for broadening the conceptualisation of postnatal distress. Exploratory factor analysis yielded 10 factors accounting for 58% of the variance: a general mental status factor (28% of the variance), psychoticism / morbid thoughts (6%), generalized anxiety (6%), panic (3%), guilt/self-criticism (3%), compulsive behaviour (3%), hyper-vigilance (2%), contentment (2%),

negative body image (2%) and manic behaviour (2%) (Marrs et al., 2009). Depressive symptoms did not load clearly onto one factor and were not the most commonly endorsed items. Internal consistency ranged from 0.67 (manic behaviour) to 0.92 (mental status). Forty-three percent of the sample had a previous mental health diagnosis and thus the factor structure may differ in more representative perinatal samples. The mental status factor captures some of the core emotional and cognitive features of anxiety and depression that are similar to the dsyphoria scale on the IDAS. This may indicate that perinatal distress is better conceptualised by an underlying factor of general distress onto which anxiety and or depressive symptoms load, rather than as two distinct factors of anxiety and depression.

Discussion

This systematic review highlights the paucity of research validating measures of anxiety in perinatal populations and the heterogeneity of studies that do. It also raises important issues concerning the choice of anxiety screening measure to use in perinatal populations. Given the limited data, recommendations for each of the measures reviewed are cautiously provided followed by key issues that have arisen from this review.

The limited number of studies validating the HADS-A makes it difficult to draw conclusions about the use of this measure in pregnant samples. Despite excellent sensitivity and specificity, low internal consistency and discrepancies in factor structure and the prevalence of probable anxiety disorder identified using the recommended cut-off of 8 are a concern, although these may be due to cultural or methodological differences. Therefore more research is required to validate the HADS-A in perinatal samples before it is possible to recommend use of this measure.

The STAI has demonstrated good criterion validity in pregnant samples although positive predictive value was low, which is an issue if the measure is to be used for screening. The six-item version may be as useful as the longer version for screening in pregnancy. In the postpartum period, discriminant validity has been demonstrated but more research is needed to demonstrate criterion validity. Factor analysis is also necessary in pregnant and postpartum samples. The overlap with depressive symptoms must be kept in mind when using the STAI. The STAI is the only measure that separates state from trait anxiety. More research is needed to investigate stability of the trait scale over time.

The GHQ may be suitable as a screen for general psychopathology in perinatal populations, with the GHQ-12 or the modified GHQ-30 offering the best sensitivity and specificity. Only the GHQ-28 can offer a specific measure of anxiety and further factor analysis is required to determine the factor structure in perinatal populations. Results from different scoring methods are mixed. If different scoring methods result in different probable cases, the clinician or researcher cannot be confident that the measure accurately identifies women experiencing distress and in need of further assistance (Jomeen and Martin, 2004; Swallow et al., 2003).

Concerning other measures, the limited evidence suggests that the DSSI/sAD is not appropriate for use in perinatal populations in its current form. More research is needed before recommending the use of the SCL-90 and HSCL-25 in perinatal populations. The use of a single item 'Do you feel very anxious?' (Sagrestano et al., 2001) reflects the current guidelines for screening for postnatal depression but is unlikely to encompass enough information to successfully screen for different types of anxiety disorder and symptoms. The Pregnancy Anxiety Scale (Levin, 1991) may be useful in this specific population but studies are needed to validate it as none have been done since its creation. The K10 may be of similar use to the GHQ-12 with the benefit of showing further power to discriminate anxiety from general distress. The IDAS could be a useful measure in perinatal populations due to the inclusion of detailed scales of symptoms of anxiety disorders and depression in the same measure. It has shown a clear factor structure in multiple samples but needs to be validated against a clinical interview in perinatal samples.

The DASS-21 has the apparent advantage of differentiating between diagnoses of anxiety, depression and stress and it is largely free of somatic items. The DASS remains to be validated against a clinical interview in perinatal samples, although research in non-perinatal samples has shown that DASS and DASS-21 scores discriminated between diagnoses of major depressive disorder, panic disorder, social anxiety disorder , obsessive compulsive disorder and specific phobia; and there is a clear three-factor structure with items loading onto the expected factors (Antony et al., 1998).

A key issue arising from this review is the lack of measures of anxiety specific to pregnant or postnatal women. To date, only one anxiety measure has been published that is specific to this population (the Pregnancy Anxiety Scale). No measures have been designed for use in the postpartum. This is in contrast to postnatal depression where a range of specific measures have been developed, such as the EPDS and the Postpartum Depression Screening Scale. This has not been so with measures of anxiety despite the overlap of physiological symptoms of anxiety and physiological symptoms occurring in pregnancy and the postpartum that may need to be disentangled.

Related to the need for specific measures, content validity of general measures has not been evidenced as many measures contain items that may not be appropriate to perinatal populations. In early pregnancy many symptoms (e.g. nausea, vomiting, aches, dizziness) may lead to inflated scores on measures that include somatic symptoms - leading to a higher number of false positives (Swallow et al., 2003). The high number of 'cases' of women identified by the GHQ supports this position. Physical symptoms in the postpartum are also likely to inflate scores, although women may feel such symptoms are acceptable and they do not contribute to psychological disorder. Inclusion of somatic symptoms is also likely to result in a different factor structure in data from perinatal populations and subscale scores may thus not be valid. Care must be taken to ensure that factor structure is determined in the population of choice before interpreting subscale scores because a subscale that purports to measure somatic distress and thus contribute to an overall ill health score may really just reflect the normal physical symptoms of pregnancy.

Thorough validation of measures is also required. Concurrent, discriminant and factorial validity are recognised as important components of ensuring a measure is valid but they do not provide a complete picture and it is important that measures are validated against a clinical interview when they are being used in a new population. The GHQ12, 28 and 30, STAI, and K-10 have been validated against a clinical interview in perinatal samples, but all other measures included remain to be validated in this way. Measures also need to be validated in pregnancy *and* in the postpartum if they are to be used in both populations. Only four studies validated measures against a gold standard in the postpartum. If the purpose of a measure is to screen for anxiety and indicate those women who need further consultation, it is vital to know that it can detect as many true cases as possible as evidenced by sensitivity and positive predictive value, which can only be calculated if the measure has been validated against a gold standard. Clinician judgement and use of the gold standard will also have an effect on the utility of validity coefficients (Joiner et al., 2005). Only the GHQ had been validated enough in perinatal populations to enable comparison of validity coefficients, but not specifically for the anxiety subscale of the 28-item version.

Another issue is whether to use general measures of distress or anxiety specific measures. The measurement context will influence this decision. In the research context it may be desirable to use a questionnaire that can measure a continuum of anxiety symptoms associated with other variables, such as the STAI. However, in clinical practice and large surveys, the general measures included in this review (GHQ, SCL-90-R, K10) may offer the best utility in detecting cases of psychiatric morbidity. However these measures will not offer as much information about the experience of anxiety as a specific anxiety measure.

Additionally, general anxiety measures that have not been designed for the perinatal period may miss facets of anxiety that are particular to this period. The Pregnancy Anxiety Scale (Levin, 1991) measures anxiety related to pregnancy, anxiety related to childbirth, and anxiety related to hospitalization. A number of measures have been developed that are specific to pregnancy and consider aspects related to anxiety such as worry or daily hassles but do not measure anxiety itself (Cambridge Worry Scale: Green et al., 2003; Pregnancy Experience Scale: DiPietro et al., 2004). These measures are not designed for the postpartum. Furthermore some women may be more prone to developing anxiety as consequence of specific physiological and psychological processes associated with birth which raise additional considerations for measurement (Heron et al., 2004). Further research and meta-analyses of prevalence of anxiety disorders in perinatal populations will assist in determining which anxiety disorders pose particular problems for perinatal populations.

A key issue arising from this review is the overlap of depression and anxiety symptoms in the perinatal period and how to differentiate them. Concurrent validity will only yield useful information in as much as the reference measure (or clinical interview) measures that which it purports to measure. Self-report measures of a construct should correlate more highly with another measure of the same construct, than with a purported measure of a different construct (Watson et al., 2008, Clark and Watson, 1991). For example, the EPDS is used as a measure of depression but evidence suggests a few items on the EPDS measure anxiety (Phillips et al., 2009; Jomeen and Martin, 2005). Thus one cannot know whether the measure being validated is correlating with anxiety or depression symptoms. In this review, both the HADS-A and the state scale of the STAI correlated more highly with the

EPDS than with other measures of depression supporting the evidence that the EPDS does detect anxiety but obscuring the extent to which the STAI and HADS-A are specific to anxiety or share overlap with symptoms of depression in perinatal samples.

More generally, the case for broadening the measurement of postnatal distress has been put forward (Heron et al., 2004; Miller et al., 2006; Ross et al., 2003). This has begun to be captured in development of measures. The IDAS, although a general symptom measure, shows evidence of a general distress scale underlying factors of anxiety and depression. This may be useful in capturing the close relationship between depression and anxiety in the perinatal period. The DASS-21 is also of interest in perinatal populations due its separation of anxiety, stress and depression and exclusion of somatic factors. Both of these measures have the advantage of encompassing a broader concept of postnatal distress.

Researchers and clinicians will need to determine the best measure for their particular needs in screening or researching anxiety in pregnancy or the postpartum, although some tentative conclusions can be drawn. The STAI is a robust and specific measure of anxiety that has been validated against clinical interview and shows discriminant and predictive validity in perinatal samples and may therefore be most useful for research purposes. The K10 may be the best short screening measure due to its ability, over the GHQ-12, to differentiate anxiety disorders. The DASS-21 shows appropriate content and a logical next step is to validate it against clinical interview. The DASS-21 and IDAS may be most useful for measuring specific but multiple types of distress in pregnancy and the postpartum.

Whilst some recommendations have been made about measures that are suitable for measuring perinatal anxiety, this review is limited by the small number of studies that have validated self-report measures of anxiety in perinatal populations. Studies vary widely on validation methods and sampling so are not easily comparable. Data is mixed for all measures of anxiety which may be due to methodological or cultural differences. The review may also be limited by the inclusion of studies in English only, and papers may have been missed due to the search strategy. A review of all measures

of anxiety used at particular perinatal times (as opposed to studies validating self-report measures) would provide useful normative data for researchers. Perinatal anxiety is an increasingly recognised concern which will need brief, feasible and accurate self-report measures to assist research and screening of women.

References

Abiodun, O.A., 1994. A validity study of the Hospital Anxiety and Depression Scale in General Hospital Units and a community sample in Nigeria. Br. J. Psychiatry. 165, 669-672

Abiodun, O.A., Adetoro, O.O., Ogunbode, O.O., 1993. Psychiatric morbidity in a pregnant population in Nigeria. Gen. Hosp. Psychiatry. 15, 125-128.

Aderibigbe, Y.A., Riley, W., Lewin, T., Gureje, O., 1996. Factor structure of the 28-item general health questionnaire in a sample of antenatal women. Int. J. Psychiatry Med. 26, 263-269.

Aderibigbe, Y.A., Gureje, O., 1992. The validity of the 28-item general health questionnaire in a Nigerian antenatal clinic. Soc. Psychiatry Psychiatr. Epidemiol. 27, 280-283.

Antony, M.M., Bieling, P.J., Cox, B.J., Enns, M.W., Swinson, R.P., 1998. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol. Assess. 10, 176-181.

Ayers, A. (2001) Assessing psychopathology in pregnancy and the postpartum. J. Psychosom. Obstet. Gynaecol. 22, 91-102.

Barnett, B., Parker, G., 1985. Professional and non-professional intervention for highly anxious primiparous mothers. Br. J. Psychiatry. 146, 287-293.

Barnett, B., Parker, G., 1986. Possible determinants, correlates and consequences of high levels of anxiety in primiparous mothers. Psychol. Med. 16, 177-185.

Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety: Psychometric properties. J. Consult. Clin. Psychol. 56, 893-897.

Bedford, A. & Foulds, G.A., 1978. Manual of the Personal Disturbance Scale (DSSI/sAD). Windsor: NFER-Nelson.

Berle, J., Mykletun, A., Daltveit, A., Rasmussen, S., Holsten, F., Dahl, A., 2005. Neonatal outcomes in offspring of women with anxiety and depression during pregnancy. A linkage study from the Nord-Trondelag Health Study (HUNT) and Medical Birth Registry of Norway. Arch. Women's Ment. Health. 8, 181-189.

Bjelland, I., Dahl, A.A., Haug, T.T., Neckelmann, D., 2002. The validity of the Hospital Anxiety and Depression Scale: an updated review. J. Psychiat. Res. 52, 69-77.

Boyd, R.C., Le, H.N., Somberg, R., 2005. Review of screening instruments for postpartum depression. Arch. Women's Ment. Health. 8, 141-153.

Brouwers, E.P.M., van Baar, A.L., Pop, V.J.M., 2001. Does the Edinburgh Postnatal Depression Scale measure anxiety? J. Psychosom. Res. 51, 659-663.

Clark, L.A., Watson, D., 1991. Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. J. Abnorm. Psychol. 100, 316-336.

Cohen, J., 1992. A power primer. Psychol. Bull. 112, 155-159.

Costello, C.G. & Comrey, A.L., 1967. Scales for measuring depression and anxiety. J. Psychol. 66, 303-313.

Cox, J.L., O'Connor, Y.M., 1983. Prospective study of the psychiatric disorders of childbirth by self-report questionnaire. J. Affect. Disord. 5, 1-7.

Cox, J.L., Holden, J.M., Sagovsky, R., 1987. Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. Br. J. Psychiatry. 150, 782-786.

Crisp, A.H., Jones, M.G., Slater, P.,1978. The Middlesex Hospital Questionnaire: A validity study. Br. J. Med. Psychol. 51, 269-280.

Derogatis, L.R., 1992. SCL-90-R Administration, Scoring and Procedures Manual-II. Towson, MD: Clinical Psychometric Research.

Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H., Covi, L., 1974. The Hopkins Symptom Checklist: A measure of primary symptom dimensions. Mod. Probl. Pharmacopsychiatry. 7, 79-110.

DiPietro, J.A., Ghera, M.M., Costigan, K., Hawkins, M., 2004. Measuring the ups and downs of pregnancy stress. J. Psychosom. Obstet. Gynaecol. 25, 189-201.

DiPietro, J., 2010. Psychological and psychophysiological considerations regarding the maternal-fetal relationship. Infant Child Dev. 19, 27-38.

Donker, T., Comijs, H., Cuijpers, P., Terluin, B., Nolen, W., Zitman, F., Penninx, B., 2010. The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. Psychiatry Res. 176, 45-50.

Field, T., Diego, M., Hernandez-Reif, M., Figueiredo, B., Deeds, O., Ascencio, A., Schanberg, S., Kuhn, C., 2010. Comorbid depression and anxiety effects on pregnancy and neonatal outcome. Infant Behav. Dev. 33, 23-29.

Furukawa, T.A., Kessler, R.C., Slade, T., Andrews, G., 2003. The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. Psychol. Med. 33, 357-362.

Geisinger, K.F., 1994. Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. Psychol. Assess. 6, 304-312.

Giakoumaki, O., Vasilaki, K., Lili, L., Skouroliakou, M., Liosis, G., 2009. The role of maternal anxiety in the early postpartum period: Screening for anxiety and depressive symptomatology in Greece. J. Psychosom. Obstet. Gynaecol. 30, 21-28.

Gibson, J., McKenzie-McHarg, K., Shakespeare, J., Price, J., Gray, R., 2009. A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. Acta Psychiatr. Scand. 119, 350-364.

Glasheen, C., Richardson, G.A., Fabio, A., 2010. A systematic review of the effects of postnatal maternal anxiety on children. Arch. Women's Ment. Health. 13, 61-74.

GL Assessment, General Health Questionnaire, Frequently asked questions, accessed 12 April 2010. http://www.gl-assessment.co.uk/health_and_psychology/resources/general_health_questionnaire/faqs.asp?css=1

Goldberg, D.P., 1972. The Detection of Psychiatric Illness by Questionnaire. (Maudsley monograph no 21). Oxford: Oxford University Press

Goldberg, D.P., Hillier, V.F., 1979. A scaled version of the General Health Questionnaire. Psychol. Med. 9, 139-145.

Goldberg, D.P., Gater, R., Sartorius, N., Ustun, T.B., Piccinelli, M., Gureje, O., Rutter, C., 1997. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol. Med. 27, 191-197.

Grant, K-A., McMahon, C., Austin, M-P., 2008. Maternal anxiety during the transition to parenthood: A prospective study. J. Affect. Disord. 108, 101-111.

Green, J.M., Kafetsios, K., Statham, H.E., Snowdon, C.M., 2003. Factor structure, validity and reliability of the Cambridge Worry Scale in a pregnant population. J. Health Psychol. 8, 753-764.

Greenhalgh, T., 2006. How to read a paper: the basics of evidence-based medicine (3rd edition). Oxford: Blackwell Publishing.

Heron, J., O'Connor, T.G., Evans, J., Golding, J., Glover, V., ALSPAC Team., 2004. The course of anxiety and depression through pregnancy and the postpartum in a community sample. J. Affect Disord. 80, 65-73.

Hundley, V., Gurney, E., Graham, W., Rennie, A.M., 1998. Can anxiety in pregnant women be measured using the State-Trait Anxiety Inventory? Midwifery. 14, 118-121.

Huizink, A.C., Mulder, E.J.H., Robles de Medina, P.G., Visser, G.H.A., Buitelaar, J.K., 2004. Is pregnancy anxiety a distinctive syndrome? Early. Hum. Dev. 79, 81-91.

Johnson, R.C., Slade, P., 2003. Obstetric complications and anxiety during pregnancy: Is there a relationship? J. Psychosom. Obstet. Gynecol. 24, 1-14.

Joiner, T.E.Jr., Walker, R.L., Pettit, J.W., Perez, M., Cukrowicz, K.C., 2005. Evidence-based assessment of depression in adults. Psychol. Assess. 17, 267-277.

Jomeen, J., Martin, C.R., 2005. Confirmation of an occluded anxiety component within the Edinburgh Postnatal Depression Scale (EPDS) during pregnancy. J. Reprod. Infant Psychol. 23, 143-154.

Jomeen, J., Martin, C.R., 2004. Is the Hospital Anxiety and Depression Scale (HADS) a reliable screening tool in early pregnancy? Psychol. Health. 19, 787-800.

Kaaya, S.F., Fawzi, M.C.S., Mbwambo, J.K., Lee, B., Msamanga, G.I., Fawzi, W., 2002. Validity of the Hopkins Symptom Checklist -25 amongst HIV-positive pregnant women in Tanzania. Acta Psychiatr. Scand. 106, 9-19.

Karimova, G.K., & Martin, C.R., 2003. A psychometric evaluation of the Hospital Anxiety and Depression Scale during pregnancy. Psychol. Health Med. 8, 89-103.

Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S-L., Walters, E.E., Zaslavsky, A.M., 2002. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol. Med. 32, 959-976.

Kitamura, T., Sugawara, M., Aoki, M., Shima, S., 1989. Validity of the Japanese version of the GHQ among antenatal clinic attendants. Psychol. Med. 19, 507-511.

Kitamura, T., Shima, S., Toda, M.A., Sugawara, M., 1993. Comparison of different scoring systems for the Japanese version of the General Health Questionnaire. Psychopathology, 26, 108-112.

Kitamura, T., Shima, S., Sugawara, M., Toda, M.A., 1994a. Temporal variation of validity of self-rating questionnaires: Repeated use of the General Health Questionnaire and Zung's Self-rating Depression Scale among women during antenatal and postnatal periods. Acta Psychiatr. Scand. 90, 446-450.

Kitamura, T., Toda, M.A., Shima, S., Sugawara, M., 1994b. Validity of the repeated GHQ among pregnant women: A study in a Japanese General Hospital. Int. J. Psychiatry Med. 24, 149-156.

Kline, P., 2000. A Psychometrics Primer. London: Free Association Books.

Levin, J.S., 1991. The factor structure of the Pregnancy Anxiety Scale. J. Health Soc. Behav. 32, 368-381.

Lobel, M., Cannella, D.L., Graham, J.E., DeVincent, C., Schneider, J., Meyer, B.A., 2008. Pregnancy-specific stress, prenatal health behaviours, and birth outcomes. Health. Psychol. 27, 604-615

Lovibond, S.H., Lovibond, P.F., 1995. Manual for the Depression Anxiety Stress Scales (2nd edition). Sydney: Psychology Foundation of Australia.

Marrs, C.R., Durette, R.T., Ferraro, D.P., Cross, C.L., 2009. Dimensions of postpartum psychiatric distress: Preliminary evidence for broadening clinical scope. J. Affect. Disord. 115, 100-111.

Marteau, T.M., Bekker, H., 1992. The development of a six-item short form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br. J. Clin. Psychol. 31, 301-306.

Martin, C.R., Jomeen, J., 2003. Is the 12-item General Health Questionnaire (GHQ-12) confounded by scoring method during pregnancy and following birth? J. Reprod. Infant Psychol. 21, 267-278.

Matthey, S., Barnett, B., Howie, P., Kavanagh, D.J., 2003. Diagnosing postpartum depression in mothers and fathers: Whatever happened to anxiety? J. Affect. Disord. 74, 139-147.

Milgrom, J., Gemmill, A.W., Bilszta, J.L., Hayes, B., Barnett, B., Brokks, J., Ericksen, J., Ellwood, D., Buist, A., 2008. Antenatal risk factors for postnatal depression: A large prospective study. J. Affect. Disord. 108, 147-157.

Miller, R.L., Pallant, J.F., Negri, L.M., 2006. Anxiety and stress in the postpartum: Is there more to postnatal distress than depression? BMC Psychiatry. 6 (12), EPub: 2006, Mar 24. doi:10.1186/1471-244X/6/12

Mirza, I., Jenkins, R., 2004. Risk factors, prevalence, and treatment of anxiety and depressive disorders in Pakistan: A systematic review. Br. Med. J. 328, 794-797.

Moss, K.M., Skouteris, H., Wertheim, E.H., Paxton, S.J., Milgrom, J., 2009. Depressive and anxiety symptoms through late pregnancy and the first year post birth: An examination of prospective relationships. Arch. Women's Ment. Health, 12, 345-349.

Muzik, M., Klier, C.M., Rosenblum, K.L., Holzinger, A., Umek, W., Katschnig, H., 2000. Are commonly used self-report inventories suitable for screening postpartum depression and anxiety disorders? Acta Psychiatr. Scand. 102, 71-73.

Myers, K., Winters, N.C., 2002. Ten-year review of rating scales. I: Overview of scale functioning, psychometric properties, and selection. J. Am. Acad. Child Adolesc. Pyschiatry. 41, 114-122.

Navarro, P., Ascaso, C., Garcia-Esteve, L, Aquado, J., Torres, A., Martin-Santos, R., 2007. Postnatal psychiatric morbidity: A validation study of the GHQ-12 and the EPDS as screening tools. Gen. Hosp. Psychiatry. 29, 1-7.

Nott, P.N., Cutts, S., 1982. Validation of the 30-item General Health Questionnaire in postpartum women. Psychol. Med. 12, 409-413.

Phillips, J., Charles, M., Sharpe, L. & Matthey, S. (2009) Validation of the subscales of the Edinburgh Postnatal Depression Scale in a sample of women with unsettled infants. Journal of Affective Disorders, 118, 101-112. doi:10.1016/j.jad.2009.02.004

Phillips, J., Sharpe, L., Matthey, S., Charles, M., 2009. Maternally focussed worry. Arch. Women's Ment. Health. 12, 409-418.

Prince, M., 2008. Measurement validity in cross-cultural comparative research. Epidemiol. Psichiatr. Soc.17, 211-220.

Ross, L.E., Gilbert Evans, S.E., Sellers, E.M., Romach, M.K., 2003. Measurement issues in postpartum depression part 1: Anxiety as a feature of postpartum depression. Arch.Women's Ment. Health. 6, 51-57.

Ross, L.E., McLean, L.M. 2006. Anxiety disorders during pregnancy and the postpartum period: A systematic review. J. Clin. Psychiatry. 67, 1285-1298.

Sagrestano, L.M., Rodriguez, A.C., Carroll, D., Bieniarz, A., Greenberg, A., Castro, L., Nuwayhid, B., 2002. A comparison of standardized measures of psychosocial variables with single-item screening measures used in an urban obstetric clinic. J. Obstet. Gynecol. Neonatal Nurs. 31, 147-155.

Sharp, D., 1988. Validation of the 30-item General Health Questionnaire in early pregnancy. Psychol. Med. 18, 503-507.

Shevlin, M., Brunsden, V., Miles, J.N.V., 1998. Alternative factor models of the Personal Disturbance Scale (DSSI/sAD). Pers. Individ. Dif. 25, 569-574.

Snaith, R.P., 2003. The Hospital Anxiety and Depression Scale. Health Qual. Life Outcomes. 1: 29. doi:10.1186/1477-7525-1-29 Accessed 19 January 2010.

Spielberger, C.D., Gorsuch, R.L., Lushene, R.E., 1970. Manual for the State-Trait Anxiety Inventory (STAI). Palo Alto, CA: Consulting Psychologists Press.

Spies, G., Stein, D.J., Roos, A., Faure, S.C., Mostert, J., Seedat, S., Vythilingum, B., 2009. Validity of the Kessler 10 (K10) in detecting DSM-IV defined mood and anxiety disorders among pregnant women. Arch.Women's Ment. Health. 12, 69-74.

Stein, G., Van den Akker, O., 1992. The retrospective diagnosis of postnatal depression by questionnaire. J. Psychosom. Res. 36, 67-75.

Stuart, S., Couser, G., Schilder, K., O'Hara, M., Gorman, L., 1998. Postpartum anxiety and depression: Onset and comorbidity in a community sample. J. Nerv. Ment. Dis. 186, 420-424.

Sutter-Dallay, A.L., Giaconne-Marcesche, V., Glatigny-Dallay, E., Verdoux, H., 2004. Women with anxiety disorders during pregnancy are at increased risk of intense postnatal depressive symptoms: A prospective survey of the MATQUID cohort. Eur. Psychiatry. 19, 459-463.

Swallow, B.L., Lindow, S.W., Masson, E.A., Hay, D.M., 2003. The use of the General Health Questionnaire (GHQ-28) to estimate prevalence of psychiatric disorder in early pregnancy. Psychol. Health Med. 8, 213-217.

Teixeira, J., Fisk, N., Glover, V., 1999. Association between maternal anxiety in pregnancy and increased uterine artery resistance index: Cohort based study. Br. Med. J. 318, 153-157

Vacha-Haase, T., Kogan, L. R., Thompson, B., 2000. Sample compositions and reliabilities in published studies versus those in test manuals: Validity of score reliability inductions. Educ. Psychol. Meas. 60, 509-522.

Van den Bergh, B.R.H, Marcoen, A., 2004. High antenatal maternal anxiety is related to ADHD symptoms, externalizing problems, and anxiety in 8 and 9 year olds. Child. Dev. 13, 1085-1097.

Van den Bergh, B.R.H., Mennes, M., Oosterlaan, J., Stevens, V., Stiers,P., Marcoen, A., Lagae, L., 2005. High antenatal maternal anxiety is related to impulsivity during performance on cognitive tasks in 14- and 15- year olds. Neurosci. Biobehav. Rev. 29, 259-269.

Watson, D., O'Hara, M.W., Chmielewski, M., McDade-Montez, E.A., Koffel, E. Naragon, K., Stuart, S., 2008. Further validation of the IDAS: Evidence of convergent, discriminant, criterion and incremental validity. Psychol. Assess. 20, 248-259.

Watson, D., O'Hara, M., Simms, L.J., Kotov, R., Chmielewski, M., McDade-Montez, E.A., Gamez, W., Stuart, S., 2007. Development and validation of the Inventory of Depression and Anxiety Symptoms. Psychol. Assess. 19, 253-268.

Wenzel, A., Haugen, E.N., Jackson, L.C., Brendle, J.R., 2005. Anxiety symptoms and disorders at eight weeks postpartum. J. Anxiety Disord. 19, 295-311.

Whiting, P., Rutjes, A.W., Reitsma, J.B., Bossuyt, P.M., Kleijnen, J., 2003. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med. Res. Methodol. 3, 25.

Woolhouse, H., Brown, S., Krastev, A., Perlen, S., Gunn, J. 2009. Seeking help for anxiety and depression after childbirth: Results of the maternal health study. Arch. Women's Ment. Health. 12, 75-83.

Zelkowitz, P., Papageorgiou, A., 2005. Maternal anxiety: An emerging prognostic factor in neonatology. Acta Paediatr. 94, 12, 1704-1705.

Zigmond, A.S., Snaith, R.P., 1983. The hospital anxiety and depression scale. Acta Psychiatr. Scand. 67, 361-370.

Table 1. Overview of studies reviewed

Article citation	(#)	Country	*n*	Sample	Ante- / Post-natal period	Self-report Instrument(s)	Interview & (Diagnostic criteria)	Validity reported	Reliability reported	Quality Score
Abiodun et al., 1993[a]	1	Nigeria	240	PC	Antenatal 28.8% 1st trimester 33.3% 2nd trimester 45.9% 3rd trimester	GHQ-30	PSE (ICD-9)	Criterion: Sens, Spec, MR	-	9
Abiodun, 1994[a]	2	Nigeria	240	PC	Antenatal 28.8% 1st trimester 33.3% 2nd trimester 45.9% 3rd trimester	GHQ-12 HADS	PSE (ICD-9)	Criterion: Sens, Spec, MR	-	6
Aderibigbe et al., 1996[b]	3	Nigeria	277	PC	Antenatal Late 2nd- early 3rd Trimester	GHQ-28	PAS (DSM-III-R)	Factorial	-	9

Aderibigbe & Gureje, 1992[b]	4	Nigeria	277	PC	Antenatal Late 2nd –early 3rd trimester	GHQ-28	PAS (DSM-III-R)	Criterion: Sens, Spec, MR, PPV, NPV Concurrent Compared 2 different scoring methods Discriminant Function Analysis	-	8
Barnett & Parker, 1985[c]	5	Australia	147	PC	Postnatal i. 3-4 days ii. <3 weeks	STAI CC-D&A BDI ENES ISSI HSPA LES	Researcher-devised semi structured	Discriminant	-	8
Barnett & Parker, 1986[c]	6	Australia	147	PC	Postnatal i. 3-4 days ii. <3 weeks	STAI CC-D&A BDI	Author-devised semi structured	Discriminant	-	10

						ENES ISSI HSPA LES				
Cox et al., 1983	7	UK	230	PC	Ante- and postnatal i. 12 weeks gestation ii.23 weeks gestation iii. 1 week postnatal iv.5 months postnatal	DSSI/sAD Author-devised visual analogue scales	Standardised Psychiatric Interview	Criterion: MR Factorial	-	6
Grant et al., 2008	8	Australia	100	R	Ante- and postnatal i.3rd trimester ii.7 months postnatal	STAI EPDS	Mini-Plus (DSM-IV)	Criterion: Sens, spec, PPV, NPV Predictive	Internal consistency	11
Hundley et al., 1998	9	UK	217	PC	Ante- and postnatal i.34 weeks gestation ii.10 days postnatal	STAI	-	-	Test-retest (Trait scale)	4
Jomeen & Martin, 2004	10	UK	101	PC	Antenatal Late 1st –early 2nd	HADS	-	Factorial Concurrent	Internal consistency	8

					trimester	EPDS				
Kaaya et al., 2002	11	Tanzania	903	R	Antenatal 8-26 weeks gestation	HSCL-25 SF-36	SCID (DSM-IV)	Factorial Concurrent	Internal consistency	10
Karimova & Martin, 2003	12	UK, Uzbekistan	100	PC	Antenatal i.12 weeks gestation ii. 34 weeks gestation	HADS	-	Factorial	Test-retest Internal consistency	8
Kitamura et al., 1989[d]	13	Japan	108	PC	Antenatal <12 weeks gestation	GHQ-30	SADS (RDC)	Criterion: Sens, Spec, MR, PPV, NPV Discriminant Function Analysis	-	8
Kitamura et al., 1993[d]	14	Japan	108	PC	Antenatal <12 weeks gestation	GHQ-30	-	Compared 4 different scoring methods of GHQ Discriminant Function Analysis	-	8
Kitamura et al., 1994a[d]	15	Japan	108	PC	Ante- and postnatal i.1st trimester ii. 3rd trimester	GHQ-30	SADS (RDC)	Criterion: Sens, spec at 4 time points	-	9

					iii.5 days postnatal iv. 1 month postnatal					
Kitamura et al., 1994b[d]	16	Japan	108	PC	Antenatal i.1st trimester ii.3rd trimester	GHQ-30	SADS (RDC)	Criterion: Sens, Spec, MR, PPV, NPV	-	9
Levin, 1991	17	USA	266	PC	Postnatal measurement of antenatal anxiety i.After birth whilst mothers still in hospital	Pregnancy Anxiety Scale (PAS)	-	Factorial	-	9
Marrs et al., 2009	18	USA	215	Internet	Postnatal <30 days	Author devised scale	-	Factorial	Internal consistency	7
Marteau & Bekker, 1992	19	UK	200	PC	Pregnancy	STAI (state) 20, 4, 6 item versions	-	Concurrent	Internal consistency	5
Martin & Jomeen, 2003	20	UK	56	R	Ante- and postnatal i.Admission to	GHQ-12	-	Compared 3 different scoring methods of GHQ	Internal consistency	6

					hospital for birth ii. Labour iii. 28 day postnatal			Discriminant		
Miller et al., 2006	21	Australia	325	R	Postnatal 6 weeks – 6 months	DASS-21 EPDS	-	Concurrent	Internal consistency	8
Muzik et al., 2000	22	Austria	50	R	Postnatal 3 or 6 months	SCL-90-R EPDS Zung SRDS	SCID (DSM-III-R)	Discriminant	-	4
Navarro et al., 2007	23	Spain	145 3	PC	Postnatal 6 weeks	GHQ-12 EPDS	SCID (DSM-IV)	Criterion: Sens, spec, NND Concurrent Discriminant	-	10
Nott & Cutts, 1982	24	UK	212	PC	Postnatal 8-14 weeks	GHQ-30	CIS	Criterion: Sens, spec, MR, PPV, NPV Concurrent	-	9

Sagrestano et al., 2001	25	USA	166	PC	Antenatal >20 weeks gestation	PSAI (single items) STAI	-	Concurrent Discriminant	Internal consistency	9
Sharp, 1988	26	UK	179	PC	Antenatal <20 weeks gestation	GHQ-30	CIS	Criterion: Sens, spec, MR, PPV	-	9
Spies et al., 2001	27	South Africa	129	PC	Antenatal <20 weeks gestation	K-10	SCID (DSM-IV)	Criterion: Sens, spec, LR+, LR-, PPV, NPV	-	9
Stuart et al., 1998	28	USA	107	C	Postnatal i. 14 weeks ii. 30 weeks	STAI BAI EPDS BDI	-	Concurrent	-	9
Swallow et al., 2003	29	UK	273	PC	Antenatal 10-13 weeks Gestation	GHQ-28	-	Compared 2 different scoring methods of GHQ	Internal consistency	7

Watson et al., 2007	30	USA	830	C	Postnatal <4 months	IDAS BAI BDI EPDS	HRSD	Concurrent Factorial	Internal consistency	8

Notes: The same alphabet letter in superscript represents the same sample.

Sample:PC = primary care e.g. women at antenatal clinic; R = subsample of a research study; C = community

Self-report instruments: CC-D&A = Trait anxiety & depression scales (Costello & Comrey, 1967); BAI = Beck Anxiety Inventory (Beck & Steer, 1991); BDI = Beck Depression Inventory (Beck, 1961); ENES = Eysenck Neuroticism & Extroversion Scale (Eysenck & Eysenck, 1964); ISSI = Interview Schedule for Social Interaction (Henderson, 1981); HSPA = Hereford Scale of Parental Attitudes (Hereford, 1963); LES = Life Event Schedule (Barnett et al, 1983); DSSI/sAD = State of Anxiety and Depression (Bedford & Foulds, 1978); STAI = State Trait Anxiety Inventory (Spielberger, 1983); HADS = Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983); EPDS = Edinburgh Postnatal Depression Scale (Cox, Holden & Sagovsky, 1987); GHQ = General Health Questionnaire; DASS-21 = Depression Anxiety Stress Scales (Lovibond & Lovibond, 1995); SCL-90-R = Symptom Checklist 90 Revised (Derogatis, 1992); Zung SRDS = Zung self-rating depression scale (Zung, 1965);PSAI = Perinatal Self-Administered Inventory (clinic devised); Conflict Tactics Scale (Straus, 1979); CES-D = Centre for Epidemiological Studies Depression Scale (Radloff, 1977); PSS = Perceived Stress Scale (Cohen, Kamarck & Mermelstein, 1983); K-10 = Kessler 10 (Kessler et al., 2002); IDAS = Inventory of Depression and Anxiety Symptoms (Watson et al., 2007)

Interview and diagnostic criteria: ICD-9 = International classification of diseases; DSM = Diagnostic and Statistical Manual of Mental Disorders; SCID = Structured Clinical Interview for DSM; SADS = Schedule for Affective Disorders and Schizophrenia (Spitzer & Endicott, 1978); RDC = Research Diagnostic Criteria; CIS = Clinical Interview Schedule (Goldberg et al., 1970); HRSD = Hamilton Rating Scale for Depression (Hamilton, 1960); PSE = Present State Examination (Wing et al., 1974); PAS = Psychiatric Assessment Schedule; SPI = Standardised Psychiatric Interview.

Validity: Sens = sensitivity; Spec = specificity; MR = misclassification rate; PPV = positive predictive value; NPV = negative predictive value; NND = number needed to diagnose; LR+ = Positive likelihood ratio; LR- = Negative likelihood ratio;

Table 2. Obtained norms for the State-Trait Anxiety Inventory

Study	**Population**	**M (SD)**	
Time of measurement			
		State	*Trait*
Barnett and Parker, 1985[a]			
< 3 weeks postpartum	High anxiety	41.2	45.1
	Moderate anxiety	32.2	32.4
	Low anxiety	24.9	23.0
Grant et al., 2008			
3rd trimester of pregnancy	Antenatal diagnosed anxiety disorder	48.38 (11.33)	52.10 (10.63)
	Antenatal no diagnosis	32.70 (9.25)	33.57 (10.09)
32 weeks postpartum	Postnatal diagnosed anxiety disorder	45.72 (12.26)	47.17 (10.76)
	Postnatal no diagnosis	33.32 (8.73)	33.24 (8.64)
Marteau and Bekker, 1992			
Pregnancy	Pregnant women, full form (20 items)	37.6 (9.1)	-
	Pregnant women, prorated from 6 items	37.1 (11.0)	-
	Pregnant women, prorated from 4 items	38.4 (12.0)	-
	Pregnant women, abnormal screen, full-form	46.4 (14.8)	-
	Pregnant women, abnormal screen, 6 items	47.7 (15.8)	-
	Pregnant women, abnormal screen, 4 items	-	-
Stuart et al., 1998			
14 weeks postpartum	Postpartum women	30.43 (10.34)	33.21 (10.03)
30 weeks postpartum	Postpartum women	31.17 (9.91)	33.02 (9.45)

[a]Subjects were allocated to groups of high, moderate or low anxiety according to self-report scores on the STAI: ≥40 =high anxiety; 32-33 = moderate anxiety; ≤25 = low anxiety.

Table 3. GHQ validity coefficients and norms

Version (paper)	**Author recommended cut-off**	**Sensitivity (%)**	**Specificity (%)**	**Positive predictive value (%)**	**Negative predictive value (%)**	**Misclassification rate**
GHQ -12						
(Abiodun, 1994)	3	83.3	81.4	-	-	18.3
(Navarro et al., 1997)	4/5	80.6	80.4	1.6 [b]	-	-
GHQ-28						
(Aderibigbe & Gureje, 1992)						
GHQ scoring	3/4	75	83	46	95	18
C-GHQ scoring	7/8	82	85	53	96	16
GHQ-30						
(Abiodun, 1993)	5	80	80.9	-	-	19.1
(Kitamura et al.) (year)						
1st trimester (1989)	7/8	83.3	71.1	37	96	26.9
3rd trimester (1994b)		38.5	82.4	25	90	-
1st day postnatal (1994a)		27.8	79.2	-	-	-

5th month postnatal (1994a)		50.0	84.3	-	-	-
(Nott & Cutts, 1982)						
Unmodified	6/7	80	89	50	97	19
Author modified	6/7	83	87	53	96	16
(Sharp, 1988)[a]	5/6	77	78	52	96	22

Note: [a]Clinical Interview Schedule cut-off point of 17/18; [b]Sampling strategy prohibited calculation of positive predictive value, therefore the authors reported number needed to diagnose, a function of specificity and sensitivity which is stable across populations.